Is a Lack of Information Undermining Sanctions Enforcement? *

Avi Dutt, Abhiroop Mukherjee, George Panayotov, Debjit Roy, Xudong Wen

January 20, 2025

Abstract

We show that the enforcement of economic sanctions is being undermined by a lack of adequate information identifying violators. This conclusion rests on three key findings. First, we document that identifying violators who are not yet officially sanctioned can be surprisingly challenging, even when they are large, slow-moving physical objects like oil tankers. Second, we show that reducing such identification difficulties can strengthen enforcement: when suspected sanctions-violating tankers were publicly disclosed in August 2023, their earnings dropped 13%, they reduced operations in Russian, Iranian, and Venezuelan waters, and they became significantly more difficult to sell in the secondary market. Third, using a dynamic model of oil shipping – quantitatively disciplined by shipping market data – we estimate that such disclosures can redirect billions of dollars annually from sanctions violators to compliant entities.

Keywords: Sanctions enforcement, Oil transport, Machine learning, Shipping, Satellite data

*Dutt: Trademo; Mukherjee, Panayotov, and Wen: Hong Kong University of Science and Technology; Roy: Indian Institute of Management Ahmedabad (IIMA). Emails: x17avi@iima.ac.in; amukherjee@ust.hk; panayotov@ust.hk; debjit@iima.ac.in; xwenam@connect.ust.hk. We thank Vishesh Vatsal and Dfy-graviti for their help in analyzing satellite imagery, and Niranjana Unnithan and Benas Zurauskas for excellent research assistance. We also thank Sumit Agarwal, Tania Babina, Utpal Bhattacharya, Briana Chang, Darwin Choi, Shan Ge, John Griffin, Minkang Gu, Harrison Hong, Yan Ji, Navin Kumar, Erica Li, Vikram Mathur, Jagmeet Makkar, S. Lakshmi Naaraayanan, John Nash, Jun Pan, Elias Papaioannou, Krishna Prasad, Robert Ready, Vivek Reshamwala, Arkodipta Sarkar, Rik Sen, Janghoon Shon, Ashok Srinivasan, Yongxiang Wang, Constantine Yannelis, Jianfeng Yu, Xiaoyun Yu, Alminas Zaldokas, Jian Zhang and seminar participants at CKGSB, FMA Asia, Hong Kong University Summer Conference, HKUST, SAIF, The Education University of Hong Kong, and Tsinghua University for various helpful comments and suggestions. We thank the UN Global Platform for access to its AIS database (UN Statistics Division, "Global AIS Data Feed.", 2020).

Economic sanctions are enforced by requiring businesses to identify and avoid entities that violate sanctions.(OFSI, 2022, OFAC, 2024). Accordingly, sanctioning authorities mandate that if any business ends up knowingly or even unknowingly engaging with a sanctions violator – e.g., lends them money, or provides insurance or transportation – then they are themselves liable for punishment. But in order to comply with such requirements, businesses must first be able to identify who the violators are.

However, businesses are finding such identification difficult, despite considerable efforts (Deloitte, 2015, KPMG, 2023). While authorities such as the U.S. Government's Office of Foreign Assets Control (OFAC) publish lists of already sanctioned entities, there are many more sanctions violators who are not yet sanctioned (and so not on any government's list). For instance, the OFAC has added fewer than 100 oil tankers to its list since January 2021, although the Congressional Research Service estimates that 1,600 tankers have carried sanctioned oil in that time. Therefore, businesses are left to identify the remaining 1500-odd violators on their own – a task many consider beyond their abilities. For instance, Daniel Tadros, Chief Operating Officer of the American Club, one of the largest shipping insurance companies in the U.S., summarized the industry's concerns in an interview with The New York Times: "It's impossible for us to know on a daily basis exactly what every ship is doing, where it's going, what it's carrying, who its owners are. I would like to think that governments have a lot more capability, manpower, resources to follow that."

In this paper, we investigate whether sanctions enforcement is being undermined because many businesses are unable to identify sanctions violators. Specifically, we address two key questions: (1) How accurately can businesses identify sanctions violators using publicly available data? (2) If identification was made easier – say by publicly disclosing a more accurate list than businesses could create on their own – would the market shift in a way that harms violators?

We answer these questions by focusing on tankers used to transport oil by sea. Not only is oil crucial to the global economy (Brancaccio, Kalouptsidi and Papageorgiou (2020, 2023); Hamilton (1983); Ready (2018)), but also oil exports constitute a significant portion of revenue for many sanctioned countries, like Russia, Iran, and Venezuela. As a result, the success of sanctions is intimately connected to restricting the ability of these countries to move their oil (Babina et al. (2023); Brown (2020); Farzanegan (2011)).

To address the first question, we take the perspective of a third party using publicly available data to predict which tankers are sanctions violators, and then check the accuracy of these predictions. While conventional wisdom might suggest that detecting the movements of large objects like oil tankers should be easy, given the available modern satellite and ship-tracking data, we show that this view is simply incorrect.

Our analysis here is made possible by a unique "ground truth" dataset, provided to us by an anonymous port agent from a Middle Eastern country. This dataset contains the identities of all oil tankers that violated Iranian sanctions during January 2021, including 33 tankers operating in the Persian Gulf under foreign (not Iranian) flags. This dataset, though limited in size, is the only one of its kind, to the best of our knowledge, that provides the necessary ground truth for evaluating models aimed at predicting sanctions evasion in shipping – a key missing component that has hitherto limited research (e.g., Wolsing et al. (2022)).

Using this dataset, we test the predictive ability of various machine learning (ML) models aimed at identifying sanctions violators. These ML models focus on factors regarded as important for this purpose by industry experts, including the occasional disappearance of tankers from tracking systems or reported unusual travel paths that suggest that tankers are falsifying their actual routes (see Figure 1). Our analysis shows that even a sophisticated third party with the ability to conduct such an exercise would correctly identify a violator only 39% of the time at best (with 95% confidence). This detection rate is not only low, but it also likely overestimates the ease of detection overall – given that the Persian Gulf (where our ground truth comes from) is a much more geographically limited area than, say, the waterways around Russia.

Given such difficulty in identifying suspect tankers, we next address our second question. Namely, we ask whether revealing the identities of suspected sanctions violators hurts these suspects, and more importantly, whether it hurts true sanctions violators and sanctioned oil exporters. An ideal experiment to answer this question might start by abruptly releasing a list of suspect tankers to the public. One can then study how this affects the vessels on such a list compared to "counterfactual" tankers – those who are not on the suspect list, but are ex-ante similar to the listed vessels. We get close to this ideal by exploiting the disclosure by Windward.ai (LON:WNWD) of a list of suspect tankers from all over the world. This list was disclosed on the London Stock Exchange Group's Refinitiv Eikon platform on August 16, 2023. While companies like Windward had been selling similar proprietary lists to specific clients even earlier, these data were prohibitively expensive for all but a handful of very large players before this event. For example, in their analysis of the sanctions advisories for the maritime industry, IHS Markit and the Association of Certified Sanctions Specialists (IHS-ACSS, 2022) reported that "For smaller banks, the platforms and specific shipping tracking software and datasets, via third-party vendors, are considered to be very expensive in terms of their cost outlay. Every transaction per dollar for managing ship screening checks would probably not make sense in terms of cost from the bank's side." Following the disclosure, the Windward module on suspect tankers became available on Refinitiv for 356 US dollars

per month, reflecting a significant cost reduction for accessing sanctions-risk information.

We start by assessing the accuracy of Windward's sanctions-risk predictions, using our "ground truth" list of actual violators. We find that 27 out of 33 of our Iran-list tankers were among Windward's high/moderate-risk tankers.[1] In comparison, our machine learning models could, at best, detect 13 of these tankers at a 95% confidence level. Beyond our exercise, another publicly available list of violators put up by the prominent advocacy organization United Against Nuclear Iran (UANI) contained only 10 of these tankers as of August 2023. This suggests that Windward's list is more accurate than alternatives available before its release.

We then examine the extent to which Windward's disclosure helps third parties reduce their false positive rate, keeping their risk of unknowingly dealing with tankers to be sanctioned in the near future at 5%. We demonstrate that augmenting a shipping signals-based machine learning model with Windward's list reduces this false positive rate from 51% to 20%. This suggests that Windward's disclosure was likely a large information shock as far as detecting sanctions-violating tankers is concerned.

We next examine the real effects of Windward's disclosure on suspect tankers. We do so using a Propensity Score Matched Difference-in-differences setting (PSM-DiD), as well as Abadie (2005)'s semiparametric DiD. In this strategy, we compare tankers disclosed as high/moderate risk by Windward with counterfactual tankers, before and after the disclosure. Counterfactual tankers are other tankers that had a similar ex-ante propensity of being classified as high/moderate risk, but were actually disclosed to be low-risk.

In particular, we examine the response of three key outcome variables to this disclosure. First, we find that the fixture rate (i.e., contractual rental rates) for tankers disclosed as high-risk dropped by 13% after the disclosure, relative to the counterfactual. This could reflect the market's hesitation in hiring these tankers given that they may be more closely investigated – and impounded if found guilty of earlier violations – disrupting shipment plans. Second, we examine whether newly disclosed suspect tankers avoid – or increasingly specialize in – routes passing close to sanctioned oil-producing countries - Russia, Iran, and Venezuela. We find that tankers disclosed as high-risk are 25% less likely to navigate within 12 nautical miles of these countries by the end of the sample, compared to other similar tankers. This could again be due to increased scrutiny on these tankers, raising their risk of getting penalized if found to engage in suspicious activity again. Third, we examine proprietary data from Drewry and find that the chance of selling a tanker drops by 37% during a year once it is flagged as risky by Windward. This can be explained by the fact

[1]This overlap also helps cross-validate the authenticity of the Iranian list provided to us. We discuss potential reasons behind Windward's accuracy in Section 4.1.

that various sanctions regimes prohibit ownership changes for violating tankers.

In the final part of the paper, we use our empirical findings to calibrate a tractable dynamic model of the oil shipping market. This calibrated model allows us to uncover the impact of Windward's disclosure on two key actors for whom data is unavailable. First, we estimate the disclosure's effect on the shipping costs paid by the buyers and sellers of sanctioned oil (henceforth, "rogue exporters"), and on the number of shadow tankers hired by them – quantities that we do not have data on. Second, we also estimate how the disclosure affected the earnings and values of actual sanctions-violating tankers. While we know each tanker's risk label provided, e.g., by Windward, we do not know which of them have actually violated sanctions in the past (beyond our Iran data).

The calibrated model suggests that the Windward disclosure resulted in rogue exporters having to pay about 2% higher fixture rates, and reduce their overall hiring of shadow tankers by 11.7%. Further, the value of tankers that actually violated sanctions dropped by 10% due to the disclosure – this amounts to around $3 million for each large 20-year-old tanker (as those often used for transporting sanctioned oil). Given that a large number of tankers (about 1500 that are not on sanctions lists, as mentioned above) might be involved in such violations, in aggregate, this is a substantial hit to violators.

Finally, the model also produces a counter-intuitive result: this information shock also reduced the value of "Good" tankers, who have never carried sanctioned oil. This happened because rogue exporters now tend to avoid tankers that have violated sanctions in the past – out of increased fear of detection – which in turn leads these tankers to try and find employment with clean exporters. This increases the tanker supply for the latter, and reduces the earnings of good tankers. This decline is not offset by the rogue exporters' higher willingness to pay for them, because the rogue market is much smaller (and shrinks further after the disclosure).

Overall, our findings underscore the need for a shift in how sanctions violators are supposed to be identified. Currently, the burden of identification falls disproportionately on businesses. A more transparent system, wherein the identities of suspected violators are disclosed by authoritative sources, may be better. It can both alleviate the identification burden on businesses, and also enhance the overall effectiveness of sanctions by increasing costs of violations.

The rest of the paper is organized as follows: Section 1 discusses related literature, Section 2 provides a brief contextual background, Section 3 presents data, methods, and results from models detecting sanctions violators, Section 4 examines the effect of Windward's sanctions-risk disclosure, Section 5 presents a model and estimates key quantities of interest within its scope, and Section 6 concludes.

1 Related Literature

Our paper is related to the literature studying the difficulty in detecting evasion of economic embargoes and sanctions. Earlier studies on the topic include Hsieh and Moretti (2006), who use the price gap between Iraqi oil and its close substitutes to show that paper trails, even including the CIA's documentation, can underestimate the extent to which Iraq used bribes to circumvent the oil embargo. Dellavigna and La Ferrara (2010) examine how the illegal arms trade undercuts the effectiveness of the embargoes to a far greater degree than would be apparent from case-by-case evidence in investigative reports by groups such as Amnesty International. They propose a method to detect potential arms embargo violations based on reactions by individual stocks. More recently, Fisman et al. (2024) document a different type of sanctions circumvention – that of entrepot trading – using the context of Ukraine's economic blockade of the anthracite-rich breakaway Donbas region. Beyond the detection of sanctioned activity, our study relates to the literature on forensic economics, aimed at detecting illegal activities of various forms and fashion (e.g., Fisman and Wang (2017); Fisman and Wei (2004, 2009), Dimmock and Gerken (2012); Duggan and Levitt (2002); Griffin and Kruger (2023); Griffin and Maturana (2016); Griffin and Shams (2018); Jacob and Levitt (2003); Marion and Muehlegger (2008)).

We relate more generally to the literature on the economics of sanctions (e.g., Early and Preble (2019); Eaton and Engers (1992, 1999); Kaplow (1990); Lacy and Niou (2004); Mulder (2022); Siddiquee and van Bergeijk (2012); Clayton et al. (2023)). We also relate specifically to the macroeconomic, political, and firm-level effects of trade restrictions with Iran (Dizaji and van Bergeijk (2013); Haidar (2017)). More recent research discusses sanctions on Russia (e.g., Itskhoki and Mukhin (2022); Mamonov and Pestova (2023)), including a few studies on how firms, banks, and other entities affected by embargoes or sanctions respond (e.g., Crozet et al. (2021); Efing et al. (2023); Huynh et al. (2023); Lastauskas et al. (2023)). Different from this literature, our paper examines the impact of providing identifying information on suspected sanctions violators, rather than the impact of the imposition of sanctions themselves.

Finally, the disclosure of risk scores aspect of our paper is related to work on sudden disclosure of information through credit registries, as studied by Liberti et al. (2016). More generally, the diffusion of information in financial markets relates to a large literature (e.g., Hong and Stein (1999); Hong et al. (2007); Hong and Yogo (2012); Peress (2014)).

2 Detecting Suspects for Sanctions Compliance: An Overview

Most sanctions regimes, e.g., that of the U.S., can generally be divided into two categories: "primary" sanctions and "secondary" sanctions. Primary sanctions aim to change the behavior of foreign state or non-state actors by targeting economic activities that involve U.S. persons or entities, while secondary sanctions target activities of non-U.S. entities. Secondary sanctions authorize OFAC or the State Department to prevent non-U.S. entities from engaging with sanctioned entities by threatening to restrict their access to the U.S. financial system and firms. These sanctions are enforced on a strict liability basis, meaning that the sanctioning authority does not need to prove wrongful intent in order to enter an enforcement action and issue a penalty (FCA, 2024, OFAC, 2024).[2]

Moreover, such sanctions-related penalties can be severe; for example, global financial institutions have in several cases agreed to pay over a billion dollars in penalties to settle charges (e.g., $8.97 billion by BNP Paribas in 2014, $1.3 billion by Société Générale in 2018, $1.1 billion by Standard Chartered and $1.3 billion by UniCredit in 2019). Finally, even if a third party transacts with a counterparty who had violated sanctions years earlier, they would still be at risk of regulatory action, due to long statutes of limitations. This is true both for US and EU sanctions (in fact, on April 24, 2024, President Biden extended the statutes of limitations for violations under U.S. sanctions laws from five to ten years).

This prospect of heavy and retrospective penalties means that sanctions compliance efforts at most banks and businesses today include investigations of counterparties' past dealings, to detect suspicious behavior that might expose them to the risk of future sanctions. For example, Stan Barstad, the CEO of the largest tanker-only company Frontline, commented that charterers are on alert for hiring tankers with suspicious activity in their history that cannot be explained (Lloyd's List, 28 Aug. 2023). Not only does this process require third parties to spend abundant resources on due diligence, but also may lead them to avoid dealing with many "at-risk" counterparties altogether, even if these counterparties were actually legitimate. In fact, even the U.S. Treasury itself has flagged this issue – its 2021 Sanctions review highlighted that "small businesses may lack the resources to bear the costs of sanctions compliance" and "uncalibrated sanctions could unnecessarily lead them to turn down business opportunities in order to avoid these costs".

The magnitude of such foregone opportunities is often substantial (Gordon, 2016). For ex-

[2]For example, when a new owner took over Eagle Shipping in 2014, they found that the former owner had violated Myanmar sanctions in 2011-2012. The new owner had to settle with the US sanctioning authorities, despite not being themselves involved in those violations (OFAC, 2020).

ample, to avoid unwittingly facilitating money transfers for sanctioned entities, some American financial institutions have already cut U.S. dollar clearing services to Western businesses that have some presence in the Russian market, even if they are there legally (S&P Global, 2022). Similarly, wary of violating rules, banks might abandon entire sectors, even if they're not sanctioned, to avoid any potential regulatory run-ins (Bloomberg, 16 Jan. 2024).

The financial media reports that such compliance costs for banks to build industrial-scale operations around sanctions compliance could represent up to 15% of a bank's total annual spend in some cases (Lloyd's List, 3 May 2022). These concerns are corroborated by our broader analysis of major US financial institutions' annual (10-K) reports. Figure 2, based on machine-reading these reports using an LLM, shows that over 75% of the 50 largest US financial institutions reported concerns about sanctions in 2023. Moreover, nearly all of them (98%) pointed out that identifying sanctions violators is a major challenge for their business.

Beyond banks and financial institutions, other businesses have also shown pronounced sensitivity to dealing with potential sanctions violators. For example, in September 2019, the U.S. sanctioned Cosco Dalian, a subsidiary of Chinese shipping giant Cosco, for transporting Iranian oil. Charterers immediately refrained from booking *any* Cosco tanker, *not just* those of the sanctioned Dalian subsidiary. That effectively removed more than 50 tankers from the global charter market overnight, causing spot rates to surge along with concerns about possible effects on oil supplies worldwide (Reuters, Sept. 28, 2019).

Ultimately, these costs of foregone opportunities and other disruptions arising from businesses trying to steer clear of any sanctions risks depend on how easy it is for them to detect the actual wrongdoers from legitimate counterparties. In this paper, we focus on the case of oil shipping to quantify this burden on third parties of identifying potential violators, given that oil tankers are a main target for sanctions enforcement (e.g., OFAC, 2019). Industry insiders estimate the percentage of tankers that might have violated sanctions but are difficult to detect at about 10% of the overall market of around 12,000-14,000 tankers.[3] This sizeable proportion of suspect tankers increases the probability of unknowingly dealing with a sanctions violator. In the following section, we quantitatively examine the difficulty in identifying sanctions-violating tankers and discuss the resultant economic impact, as implied by our estimates.

[3] "A mysterious fleet is helping Russia ship oil around the world. And it's growing", CNN, Mar. 2023; "Dark fleet of tankers now comprises 10% of seaborne oil transport", Lloyd's List, Mar 2023; "'Shadow Fleet' of tankers hauling Russian oil swells to 600 ships, Trafigura says", Bloomberg, Feb. 2023

3 How difficult is it to detect sanctions-evading tankers?

To quantify the difficulty in detecting sanctions evaders, we take the view of a responsible entity making its best effort to comply with sanctions. We assume that such an entity employs sophisticated analytical tools like machine learning and has access to radio-signal-based ship-tracking data from the Automatic Identification System (AIS) and satellite images. We first assess the detection accuracy only based on AIS data, and then add satellite data to the model to evaluate how much it improves accuracy.

3.1 Data

We use three key datasets here: AIS data on high-frequency shipping signals, Satellite data from Sentinel-1, and our proprietary Iran dataset to assess detection performance.

The AIS dataset was originally developed for collision avoidance, and has also been used in academic research, e.g., by Brancaccio et al. (2020). AIS transmits at short intervals (1) fixed information including the vessel's International Maritime Organisation (IMO) number, Maritime Mobile Service Identity (MMSI) number, and vessel type and size, (2) dynamic information such as vessel position and speed, and (3) voyage information such as draught, destination and expected time of arrival. Dynamic information, e.g., position and speed, is updated automatically from the ship sensors, while voyage information is entered manually by the vessel operator. Vessels above 300 gross tonnage are required to carry AIS equipment and share AIS data with other concerned parties to aid navigation (IMO, 2000). AIS is a well-recognized tool for sanctions compliance (e.g., U.N. Doc. S/2019/171, 5 March 2019, Kilpatrick (2022)). Even flag state registries have de-flagged ships based on suspicions raised by AIS data (see Lloyd's List, Feb. 2020, Oct. 2020).

Second, we use data from the Sentinel-1 mission, which has two satellites that can take an image of the same location on Earth at six-day intervals. It generates images by synthetic aperture radar (SAR) and makes freely available high-resolution, all-weather Earth observations.[4] We use all satellite images that contain some parts of the Persian Gulf and were taken in January 2021. Figure 3 shows three of these images, together with enlargements that make visible individual ships at sea. The processing of satellite images was done by Dfy Graviti – a company specializing in aerospace and maritime AI, so that we can identify ships at sea from these images, and obtain their exact locations, i.e., latitude and longitude. Regulators and others recognize the usefulness of commercial satellite imagery for such identification

[4]This dataset is widely regarded as one of the most popular satellite sources of imagery currently available and, with its 5-day revisit rates, the closest one can get to free, real-time satellite observations (see https://skywatch.com/free-sources-of-satellite-data/).

(OFAC, 2023, Economist, 16 Apr. 2022).

Our third dataset comes from Iran. Testing detection ability requires knowing the 'truth' for some subset of the data. We test how successful these AIS-based predictors can be in detecting oil tankers based on our dataset of oil tankers that were seen in Iranian waters in January 2021, violating sanctions. This dataset of tankers was collected by an anonymous source with decades of experience with the Middle-Eastern shipping supply chain. It lists 33 tankers that were neither Iranian-flagged nor part of the Iran-Russia Caspian Sea trade, i.e., they were observed in Iran somewhere along the Persian Gulf coastline or territorial waters). Our data does not list specifically which Iranian port or location these tankers were at. Our source collected this data through the aggregation of a large number of text messages from contacts at ports, brokers, and charterers. We later cross-validate the accuracy of this dataset by comparing it with Windward's high-risk list.

3.2 Patterns of Suspicious Activity

According to regulators' guidance (e.g., OFAC, 2023), shipping industry actors are advised to pay specific attention to suspicious gaps in AIS transmission ("Dark activity"), deviations in routes without legitimate reasons, engaging in ship-to-ship transfers, and patterns of shipping trajectories that appear anomalous (e.g., impossibly sharp turns at high speed, see Windward: "Going dark is so 2019", August 2020). The latter is suggestive of geolocation "Spoofing" – wherein AIS data is manipulated to show a ship in one location while it is physically present in another, as if using a VPN. Such spoofing was reported by the United Nations in 2019 (U.N. Doc. S/2019/171, 5 March 2019, pages 9-11) and has spread quickly since (New York Times, 3 Sep. 2022, Economist, 16 Apr. 2022). Figure 1 illustrates episodes of spoofing and dark activity for two tankers on our dataset of Iranian sanctions violators. It shows the tankers' trajectories, as per their AIS signals. In one case, the tanker's uninterrupted AIS signals placed it in the northern part of the Persian Gulf at a time when it was observed at an Iranian port. In the second case, the tanker stopped transmitting AIS signals for four days, during which it was observed at an Iranian port.

3.3 Detection: Predictor Variables

3.3.1 Predictors using AIS Data

Following prior studies, and consistent with guidance by regulators (e.g., OFAC, 2020) and other industry experts (e.g., Wolsing et al. (2022)), we construct six categories of sanctions evasion predictors. The flexibility of machine learning allows us to incorporate various ver-

sions of these variables and combine them in a non-linear fashion. Here we list six broad categories under which our predictors fall; detailed variable descriptions are in Internet Appendix C.

- Identity Change: equals one if a tanker has changed its MMSI number or flag in the preceding six months.
- Risky Flag: equals one if a tanker is registered in a country whose flags are often linked to previous violations.
- Ship-to-ship Transfer (STS): This is a key deceptive shipping practice where a tanker loads oil from another vessel away from ports to avoid scrutiny.
- Irregular Trajectory: equals one if a tanker's trajectory in AIS shows abnormal patterns often associated with spoofing. Figure 4 illustrates an irregular trajectory from a tanker in the Persian Gulf in January 2021, which we confirm to be a spoofer based on our proprietary data.
- DBSCAN Outlier: equals one if a tanker sent signals from anomalous locations, which are obtained using a data-driven clustering method called DBSCAN.
- Dark Activity: equals one if there is an unusually long time gap between two AIS signals, as illustrated above in the bottom panel of Figure 1.

3.3.2 Predictors using Sentinel Satellite Data

Ideally, we would detect spoofing if a satellite image taken at the exact time of a tanker's AIS signal shows no ship at the location transmitted by this signal (or, shows the tanker at a location different from the transmitted one). Similarly, we would detect dark activity if we could track a tanker's trajectory during a dark period (e.g., we may find violators headed towards an Iranian port).[5]

However, this ideal situation would require frequent, if not continuous, satellite observations of multiple locations in the Persian Gulf, whereas, in fact, such observations are quite sparse – the satellites generating our data revisit each area only at six-day intervals.

To deal with this challenge of not having AIS signals on a tanker at the exact time when it is photographed by Sentinel, we interpolate tanker trajectories as follows. Say, we have

[5]Note that alternative approaches like identifying ships at port are also infeasible, given that multiple vessels of similar size and shape make it impossible to distinguish one vessel from another. Furthermore, sanctions violators have been reported to intentionally modify visual characteristics of their vessels, like repainting in a different color or changing the vessel name or IMO number on its hull (IHS-ACSS, 2022).

a Sentinel image of some region at 10 am on a given day, but the closest AIS signals we have from our target tanker were emitted at 7 am and 1 pm. Then we use the available trajectories of other tankers that started roughly where the tanker was at 7 am, and ended up where it was at 1 pm to figure out the tanker's most likely location at 10 am. Figure 5 gives an example of constructing the search area via interpolation and detailed descriptions are in the Internet Appendix C. If, after processing the image, we do not find any tanker in this search area, we conclude that the tanker was likely involved in spoofing.

To detect dark activity, we focus on tankers that go dark and then resume emitting AIS signals at about the same place and define search areas around these places – if a satellite image of such a search area that was taken within the dark period does not show any tanker there, we conclude that the respective tanker was likely involved in wrongdoing during its dark period.

We construct the following satellite-based predictors and add them to the previous list of AIS-based predictors.

- Satellite Detection: (a) For spoofing detection, the predictor equals one for day t if a tanker is not seen in its corresponding search area in $[t-r, t+r]$ where $r = 0$, 1, or 2. (b) For dark activity detection, the predictor equals one if a tanker is not seen in its corresponding search area during its dark period.

In total, we obtain 14 predictors for spoofing detection and 14 predictors for dark activity detection, in the above seven categories. Summary statistics on these predictors are shown in Table 1.

3.4 Detection: Methods

We test the ability of the above predictors to detect sanctions violators based on (1) emission of falsified AIS signals (i.e., "spoofing") or (2) turning off their AIS signals (i.e., "going dark"). First, we construct a sample at the tanker-day level to detect tanker spoofing. The dependent variable equals one when a tanker is observed in Iran, as per our proprietary dataset, while its AIS signals show it at a different location; and zero on other days. Second, we construct a sample to detect tankers going dark, i.e., not sending an AIS signal for more than 24 hours (a tanker may have multiple such events during a month). When detecting dark activity, the dependent variable equals one if a tanker has been observed in Iran during a dark period, and zero otherwise.

To combine our predictors listed in the previous section in an optimal (and potentially non-linear) way, we employ two popular machine-learning tools: decision trees and neural

networks. We use 10-fold cross-validation to determine the optimal values of model parameters and to construct model fit measures. In this procedure, we randomly divide the sample into 10 equal sub-samples, and train models with different combinations of hyper-parameter values using nine of these sub-samples. The sub-sample left out in each case is used for validating the model predictions. We repeat this procedure 10 times, and select the combination of hyper-parameters that yield the best validation performance, based on McFadden's pseudo-R^2.

To evaluate detection performance, we consider two measures: the pseudo-R^2 measure above, corresponding to the best cross-validated model, as well as a more intuitive measure derived from it, the implied false positive rate. In our context, the false positive rate indicates the proportion of business from sanctions-compliant tankers that a shipping sector firm must forgo to maintain the probability of engaging with a violator below a pre-specified level. Further details are in Internet Appendix B.

3.5 Results

Results from these machine learning models are presented in Table 2. Panel A of this table shows that even the best AIS-based detection models yield Pseudo-R^2 around 10%, for the average of detecting spoofing and dark activity. The implied false positive rates for a firm to ensure that it avoids 95% of true violators are also staggeringly high at about 50%. This implies that to maintain the risk of unknowingly doing business with a violator at 5%, a firm has to avoid more than half of its potential counterparties.

In Panel B, we enhance our sanctions detection models by adding satellite data. Adding satellite images to models improves our detection performance compared to using only AIS-based predictors, but this improvement is still modest. For example, when detecting spoofing, the Pseudo-R^2 increases from 6.7% to 11.8% for trees after including satellite images, and from 7.2% to 13.8% for neural networks. The Pseudo-R^2 for the detection of dark activity also increases from 11.5-13.3% to around 18.7%. The implied false positive rates also decline, but not very significantly. For example, to maintain a 95% probability of avoiding sanctions violators, a firm in the shipping sector has to give up 35.8-38% of business from compliant tankers.

In panel C, we examine the incremental contribution to predictive power, as well as to the false positive rate, for each category of our underlying predictors. Given the low overall Pseudo-R^2-s above, it is not surprising that incremental Pseudo-R^2-s are all low; among them, the Dark Activity and Ship-to-ship Transfer predictors seem somewhat important. At the same time, we see that for these two variables, the incremental false positive rates are

24.5% and 14.8%, respectively. This implies that relative to a model without, say, our Dark Activity predictors, one with remaining predictors would incrementally lead to a firm falsely detecting 20% of their compliant clients as potential violators.

We conclude that despite combining many of the predictors mentioned in industry and regulatory sources – and doing so in a relatively sophisticated way – detecting sanctions evaders based on AIS data is ultimately highly inaccurate, which in turn can have a significant economic impact on sanctions-complying firms. Moreover, because our ground truth dataset comes only from Iran, we restrict our analysis to the Persian Gulf region – given the geographically limited area of the Gulf, this should make our predictive exercise easier than if, for example, one were to do the same for a country as large as Russia.

3.6 Prediction accuracy and detection difficulty

At this point, one may wonder whether our results above are sufficient to assert that detection is generally difficult; after all, the poor performance of our models could just reflect limitations in either our data or our modelling capabilities.

Delving deeper into our data, we first find that predictors of sanctions-evading activity are associated not only with violating tankers, but are almost equally prevalent for compliant ones. For example, the top plots in Figure 6 show how compliant tankers (top-right) can have similarly irregular trajectories as a sanctions-violating tanker (top-left). In parallel, the bottom plot in Figure 6 shows that the distributions of dark time for violators and compliant tankers (conditional on a tanker going dark) overlap to a large extent. The main issue is that tankers may move somewhat irregularly or show periods of dark activity, etc., for various legitimate reasons (e.g., weather conditions, mechanical failures, ship traffic congestion, data issues in AIS itself, etc., see Weng et al. (2022)). Coupled with the fact that publicly available satellite data is sparse, this results in a low signal-to-noise ratio.

Second, we show in Internet Appendix B that even doubling predictive accuracy relative to our models, perhaps by somehow gaining access to significantly better data, may not be enough. A firm that could do so would still need to give up about 20% of their current clients in order to maintain sanctions-risk exposure at 5%.

Third, the difficulty in detecting evaders extends well beyond our data or methods. In order to see this, one can check how many of the 33 tankers on our ground truth dataset of Iran-sanctions violators appeared on alternative publicly available lists aimed at identifying such violators. One such example is United Against Nuclear Iran (UANI)'s "The Ghost Armada" list of suspected sanctions-violating tankers.[6] Figure 7 shows that even as of April

[6]UANI is a bi-partisan, non-profit organization in the US; they started publishing their list in November 2020, and update it periodically.

2024, UANI had detected only 13 of these tankers.

Overall, the detection of sanctions-evading tankers does indeed seem to be difficult for third parties. Of course, Governments and specific corporations (e.g., Starlink) with access to better data (possibly from many satellites) do not face such issues, but their data is not publicly accessible to most third parties who are expected to comply nonetheless.

4 Public disclosure of suspect tankers

If the difficulty in detecting sanctions evaders indeed poses a key hurdle to the effectiveness of sanctions, then one should observe two crucial features in the data: (1) it is indeed difficult to detect evaders through honest and reasonable effort; and (2) if this detection problem can somehow be alleviated – e.g., by providing a list of suspects detected with reasonable accuracy – then players who previously engaged with these suspects due to a lack of knowledge should adjust their behavior. We have already demonstrated (1) in the previous section; in this section, we shed light on (2).

4.1 The disclosure event

The event we consider as a shock to detecting sanctions violators is Windward.AI's disclosure of Sanctions Risk Labels for oil tankers on the Refinitiv Eikon information platform in August 2023. This list covers the entire global fleet of oceanic tankers and classifies the tankers' riskiness into "Low", "Moderate", "High", and "Sanctioned". It also shows the reasons for the risk classification provided, which include ID or location manipulation, flag hopping, dark activity, STS activity, etc. (A screenshot is provided in the Internet Appendix F.)

Windward uses over 100 million data points daily to map events across the world's oceans. These data include not only freely available sources, but also proprietary satellite imagery from Planet.com, weather data from Meteomatics, etc. Most importantly, Windward augments these datasets with information provided by its clients – which include the United Nations, the European Border and Coast Guard Agency (Frontex), and US agencies such as the Drug Enforcement Administration and the Office of Naval Intelligence (See Reuters, 17 Mar. 2016, Wired, 18 Mar. 2020). Observers have pointed out that Windward's product is effectively military-style signals intelligence (SIGINT) for commercial use (RAND, 2017).

Before the August 2023 event that we consider, several AI companies like Windward itself, Lloyd's List, Planet Labs, Pole Star, Spire, TankerTrackers, etc., provided similar risk lists to specific clients. However, the cost of obtaining such data was prohibitive for most market participants except the largest of players, as discussed in the introduction. After

August 2023, Windward's data became easily affordable for a subscription fee for Refinitiv Eikon of 280 GBP per month.[7]

Moreover, such a release of information impacts higher-order beliefs, as the riskiness of each tanker becomes common knowledge (Aumann (1976)). That is, even if any particular business knew the list of suspects before August (e.g., because they had paid for that information), they would now realize that everyone else is also aware of the same list. Those of them who thought earlier that it might be okay to do business with a suspect – perhaps assuming that few others knew the contents of the list – would now have to rethink. As a result, the disclosure is likely to not only change the behavior of those who previously lacked access to the suspect list, but also affect the behavior of those who had it due to this common knowledge aspect.

4.2 Is this a sizeable information shock?

The disclosure of the Windward's list of suspect tankers could have real effects on how businesses behave only if: (1) a substantial number of businesses knew about this disclosure and paid attention to it; and (2) the list itself brought new and accurate information on suspected violators.

First, we find (as shown in Figure A-2 in the Internet Appendix) that attention to Windward – as measured by page views of the Windward.AI homepage – increased dramatically right after the list was disclosed, suggesting that market participants likely paid attention to this disclosure.

Second, in order to truly bring new information about suspected evaders to the market, Windward's list must be significantly more accurate than alternative detection methods. To assess this, we benchmark its accuracy against our methods from the previous section, using our proprietary Iranian port data on actual sanctions evaders as the ground truth. Table 3 shows that out of the 33 tankers that have violated Iranian sanctions as per our dataset, 27 have been detected by Windward as risky, i.e., they are labeled high/moderate risk. The table then shows that our ML models fall significantly behind Windward's detection success rate. Even at the 90% confidence level, our models can detect at best only half the violators, compared to 27 out of 33 for Windward (with the caveat that our numbers may not be

[7]Windward continues to sell tailored products to commercial and government clients; if anything its overall sales increased from winning new contracts of this type following the disclosure of their sanctions risk data. See, for example, https://www.proactiveinvestors.co.uk/companies/news/1029531/windward-shares-swell-on-ai-maritime-contract-wins-1029531.html. This was perhaps one reason why Windward made its information available to the market at very little cost on a prominent platform like Refinitiv – their line of business identifying sanctions violators served better to demonstrate what their capabilities are to potential clients with other information needs than as an independent profitable enterprise.

directly comparable, as Windward might have added tankers violating sanctions in January 2021 to their list based on activity in months other than January 2021 itself). Further, as discussed before, another prominent public data source on sanctions-violating tankers, UANI's "The Ghost Armada" list, featured only 10 out of the 33 actual violators at the time of the Windward disclosure. These comparisons demonstrate the relative accuracy of Windward's risk evaluations.

4.3 Effect of the disclosure on the difficulty of identifying suspect tankers

We first examine whether the disclosure of a public list of suspect ships indeed makes it easier for third parties to identify whom to avoid. To understand whether, and by how much, Windward's disclosure can help, we adopt the point of view of a third party trying to predict, out-of-sample, which tankers will be sanctioned in the future. Recall that it is a key concern for businesses to be able to detect – and subsequently avoid – such tankers, given the retrospective nature of sanctions as well as their strict legal liability.

We study tankers that were actually sanctioned between August 2023 and April 2024. Using this list of tankers, we contrast the predictive ability of public data vs. Windward's list. We start with a sample of 13,997 tankers from Refinitiv (the sample filtering process is shown in the Internet Appendix Table A-2). For Windward, we take the 3,500 tankers classified as high-/moderate-risk at the end of July 2023 (i.e., right before Windward's disclosure) and contrast them with low-risk tankers.[8]

For comparison, we estimate the probability that a tanker might be sanctioned in the near future based only on publicly available data. To do so, we follow a standard training-testing procedure. We construct a training sample using data up to January 2023, where the dependent variable takes a value of one for tankers that were sanctioned between August 2022 and July 2023, i.e., six months before and after. We include predictors described in Section 3.3, and also a zero-one predictor derived from the UANI list (as described in Section 3.6).[9] We train ML models (decision trees and neural networks) as before. The testing procedure is to estimate, standing at the end of July 2023, the probability that a tanker will be sanctioned in the future.

One simple way of contrasting Windward's predictive ability with our best model using public data is to sort all tankers based on the estimated probabilities of sanctions based on

[8] In this particular exercise, we focus on all tankers to increase statistical power. Results are similar if we use only oil tankers.

[9] Besides the Persian Gulf, we extend our satellite-based predictors in this test to cover the Black Sea and the Baltic Sea to predict Russian sanctions violators.

our model, and take the top 3,500 tankers from this list. Recall that there are 3,500 tankers on Windward's risk list as of that same time point. We find that out of the 69 tankers that were actually sanctioned after July 2023, our best model using public data predicts 33, while Windward predicts 65. This comparison suggests that Windward's disclosure significantly reduces the difficulty of detecting tankers that are subsequently sanctioned.

For further comparison of out-of-sample predictability, we evaluate the performance for weighted averages of Windward's and public-data-based predicted outcomes that use various weights. This exercise is aimed at understanding whether and how much incremental predictive power Windward adds above public-data-based machine learning, and vice-versa.

Figure 8 plots the out-of-sample performance measures (i.e., pseudo-R^2 or implied false positive rate) against the corresponding weights. It shows that the prediction based on purely public data yields an out-of-sample pseudo-R^2 of 7.5% and an implied false positive rate of 51%.[10] In contrast, the prediction based only on Windward's list yields a significantly higher pseudo-R^2 of 17.2% and a lower implied false positive rate of 26%.

Interestingly, a combination of $0.58 \times$ public data $+ 0.42 \times$ Windward's list, achieves an optimal pseudo-R^2 of 21% and implied false positive rate of 20%. This suggests that machine learning on public data (of the type we described above) still has incremental value, adding to the predictability from Windward's list.

4.4 Real effects of the disclosure on suspect tankers

Next, we examine whether there are any real effects of the disclosure on suspect ships – key to the process of enforcing sanctions. In the following sections, we investigate this issue by examining how Windward's disclosure impacts the shipping sector from different aspects: tankers' freight rates (fixtures), routes, ownership changes, and identities of their charterers.

4.4.1 Estimation Methods

Suppose we want to examine the effect of disclosing the list of suspect tankers on their fixtures (freight rates). We can do so by estimating the change in the fixture of a tanker classified as high/moderate risk in Windward's list after disclosure and comparing it to the change in the fixture of a counterfactual tanker - i.e., one that is otherwise identical to the treated tanker but is not itself labeled as high/moderate risk.

[10] The implied false positive rates here are not directly comparable to those in Table 2 for two reasons: (1) the scopes are different, as this analysis encompasses all sanctions regimes, while Table 2 is limited to Iranian sanctions; and (2) the evaluation methods differ, with this analysis using out-of-sample measures, whereas Table 2 relies on cross-validation.

The main challenge for such a strategy lies in the construction of the counterfactual. This is because untreated tankers, i.e., those classified as low-risk by Windward, may not provide accurate counterfactuals if their fixtures could have evolved differently from those treated even without Windward's disclosure.

We use two methods to address this issue. The first is propensity score matching (PSM), introduced by Rosenbaum and Rubin (1983). PSM is an intuitive method and makes it easy to visualize the data using standard DiD plots. Second, we follow Abadie (2005) and employ a semiparametric DiD estimator, which is derived under a weaker identification assumption than that of PSM (see, e.g., Heckman et al. (1997), and Abadie (2005)). Further details are in Internet Appendix E.

4.4.2 Estimated Effects

A. Freight rate (Fixture) changes

We first use data from Refinitiv Eikon to estimate the treatment effect on tankers' fixtures (contractually agreed upon rental price for a tanker, as negotiated between a shipowner and a charterer). A key feature of fixtures is that they are measured in Worldscale standardized units that allow for easy comparison across different contracts; Internet Appendix F reports further details. We use the 6-month period preceding the disclosure (February to July 2023) as the pre-period, and August 2023 to March 2024 as the post-period.

To estimate our effects, we first calculate propensity scores by regressing the high/moderate-risk indicator on a set of predictors. Here we use the same methods and set of variables that we used to detect sanctions evaders in Section 3, except that: (1) we construct variables using only pre-period information, i.e., February-July 2023, and extend the scope to all tankers rather than only tankers that were in the Persian Gulf as in Section 3; (2) since our satellite measures only cover the Persian Gulf in Jan 2021, we don't include them when calculating propensity scores; (3) since here the pre-period includes multiple months, we use count variables as predictors, for example, log of one plus the number of dark events during pre-period. The summary statistics of the sample and balance tests are shown in Table 4.

After estimating the propensity score, we match each fixture contract of a treated tanker to the contracts of untreated tankers by requiring that (1) these contracts are in the same month; and (2) the matched tankers are of the same type (i.e., Medium, Panamax, Aframax, Suezmax, or VLCC). The matching weights are determined by the difference between two propensity scores with a Gaussian kernel. We report results for bandwidth parameters 0.01, 0.03, and 0.05. We first estimate the time series of treatment effects on fixture rates for high-risk tankers only, i.e., we drop moderate-risk tankers from the sample, conduct matching,

and estimate the following regression:

$$fixture_{c,i,k,t} = \left[\sum_{l=-6}^{7} \beta_l^H \times high_risk_i \times \mathbb{I}_{\{t=l\}}\right] + \alpha_i + \gamma_{k,t} + \epsilon_{c,i,k,t}, \tag{1}$$

where $fixture_{c,i,k,t}$ is the WS rate for contract c, tanker i (with tanker type k) in month t; $high_risk_i$ equals one for tankers labeled as high-risk in Windward's list, and zero otherwise; α_i is tanker fixed effects, and $\gamma_{k,t}$ is time×tanker-type fixed effects. We are interested in the average treatment effect on the treated (ATT) for high-risk tankers, given by the coefficient series $\{\beta_l^H\}_{l=-6}^{7}$.

Figure 9 shows that the PSM-DiD estimated coefficients in the pre-period $\{\beta_l^H\}_{l=-6}^{-1}$ are insignificant, indicating a lack of pre-trends. After Windward's disclosure, there is an immediate drop of around 20 WS units in fixture rates in August for high-risk tankers compared to the matched control group, which persists until the end of our sample, and is robust across specifications.

To understand what such a drop in fixtures means for tanker earnings, we can perform the following back-of-the-envelope calculation. Given the average fixture for high-risk tankers in July 2023 is 126.8 WS units as in Table 4, the 20-unit drop in August implies revenue from chartering decreases by 15.8% (20/126.8). Note that the drop reflects the market's response to the increased risk of suspect tankers having violated sanctions, not to violations themselves.

While post-period coefficients estimated in Figure 9 lie below the pre-period ones, month-by-month estimation is noisy. To mitigate this noise, we aggregate the indicators in the post-period months into a single dummy, $\mathbb{I}_{\{t\geq 0\}}$. We separately estimate ATT for high-risk and moderate-risk tankers. In a matched sample with high- and low-risk tankers, we run the following regression:

$$fixture_{c,i,k,t} = \beta^H \times high_risk_i \times \mathbb{I}_{\{t\geq 0\}} + \alpha_i + \gamma_{k,t} + \epsilon_{c,i,k,t}, \tag{2}$$

where β^H is the ATT for high-risk tankers. We conduct a similar estimation for moderate-risk tankers by replacing $high_risk_i$ with $moderate_risk_i$. Panel A in Table 5 shows the regression results from the above PSM-DiD.

Results from our second estimation strategy, i.e., Abadie (2005)'s semiparametric DiD, are presented in panel B of Table 5. Note that the number of observations in panel B is smaller than in panel A, because Abadie (2005)'s estimation is conducted on a tanker-level sample instead of a tanker-month level one for PSM-DiD. Also, to calculate the change in fixtures as the dependent variable, i.e., post-period average fixture minus pre-period average

fixture, tankers without both pre- and post-fixtures will be dropped in panel B.

The results show that, after the information shock, the fixtures of high-risk tankers decreased by an average of 16.45 WS units across specifications, with a range of 13.75-20.71 units. This translates to a drop in earnings of around 13% (=16.45/126.8), relative to the counterfactual. We do not find any significant change in the fixtures for tankers classified as being at moderate risk by Windward.

Finally, note that these results are not only robust to methods and bandwidth choice across specifications, but also remain similar if we use propensity scores from a logistic regression (rather than the baseline machine learning models), drop August 2023 (to allow about two weeks for the disclosure effect to come to force), drop Russia (to show that our effects are not coming solely from Russia-related changes in maritime activity around this time), and even if we bootstrap standard errors. These results are presented in Table A.2 in the Internet Appendix.

B. Tanker route changes

Here we examine the effect of the information shock on tanker routes from AIS, with specific reference to Iran, Venezuela (only before 2024, because the U.S. lifted sanctions on Venezuela in 2024), and Russia. We define the relevant boundaries for these sanctioned countries using their territorial seas, which extend to 12 nautical miles (nm) from the coast of a country (we get this data from the Flanders Marine Institute).

To estimate the effects of the disclosure, the dependent variable is a zero-one indicator of AIS signals being emitted within the territorial seas of sanctioned countries each month. Our test design here mirrors that for the fixtures from the preceding section. Figure 10 plots the estimated coefficients $\{\beta_l^H\}_{l=-6}^{-1}$, which can be interpreted as the probability of passing near sanctioned countries. The coefficients in the pre-period are all insignificant at the 95% confidence level, which suggests a lack of pre-trends. In the post-period, there is a gradual drop that reaches around 10 percentage points at the end of our sample.

In Table 6, we present the estimation of ATT on the route change by combining post-period months into a single dummy and estimating separately for high-risk and moderate-risk tankers, again as before. As tanker AIS signals are much more frequent than tanker fixtures, we have more observations for this route test than the previous fixture test.

The results using PSM-DiD in panel A show that the high-risk tankers are around seven percentage points less likely to pass within 12 nm of the sanctioned countries, while the effect on moderate-risk tankers is insignificant. For the magnitude of the effect, note that the average probability of showing signals in those areas is 40 percentage points for high-risk tankers in July, which corresponds to a 17% (6.8/40) decrease for the average effect in the

post-period. According to Figure 10, this magnitude is larger towards the end of our sample, e.g., 25% (10/40). In panel B, the estimation using Abadie (2005)'s semiparametric DiD gives similar results.

Overall, our evidence suggests that high-risk tankers avoid getting close to the sanctioned countries after Windward's disclosure.

C. Tanker ownership changes

In this subsection, we examine how conditions in the market for used tankers change after the information shock. We might expect high/moderate-risk tankers to become more difficult to sell in the used tankers market, since various sanctioning regimes explicitly prohibit transactions involving violating entities. For example, Council Regulation (EU) 833/2014 prohibits EU nationals, residents, or entities from directly or indirectly selling or otherwise transferring ownership of tankers falling under Harmonized System (HS) Code 8901 20 for the transport of crude oil or petroleum products listed in Annex XXV of this regulation (Skadden, 2024).

We use a proprietary set of tanker ownership data (from Drewry), which shows the tankers' owner name and country in six snapshots – as of December 2022, March, July, September, and December 2023, and March 2024. We hypothesize that after Windward's list is released, high/moderate-risk tankers become less liquid as shipowners are reluctant to buy them. Note that the time spans between consecutive snapshots are different; so we make them comparable by annualizing the ownership change variable within each snapshot. Again, we use both PSM-DiD and Abadie (2005)'s semiparametric DiD to estimate the treatment effect, as before. The dependent variable here equals one if the owner names are different across two snapshots and zero otherwise.

The results are shown in Figure 11 and in Table 7. Panel A of that table shows results from PSM-DiD estimation which indicate that after the information shock, high-risk (moderate-risk) tankers are 16.4-18.3 (15.2-16.8) percentage points less likely to change owners. Relative to a pre-disclosure annualized turnover of 45.4 percentage points for high-risk tankers, this is a 37% (17/45.4) drop in liquidity. In panel B, the results from Abadie's DiD suggest a similarly significant drop in the probability of tanker owner change. Overall, this suggests that the risk disclosure from Windward makes high/moderate-risk tankers less desirable, reducing their turnover.

D. Charterer changes

In this subsection, we examine whether U.S. or U.S.-allied companies stop chartering tankers

classified as high/moderate risk by Windward. The charterers' names are provided in the fixture data and we manually search these companies' headquarters via LinkedIn or their websites to determine their countries. We define the U.S.-allied charterers as companies from the UK, European Union, Australia, New Zealand, or Japan. The dependent variable here is at the charterer-month level – it is the number of tankers of a given risk type that each charterer employs in a month. The regression specification is

$$n_tankers_{h,t} = \left[\beta^U \times US_charterer_h + \beta^A \times US_allied_charterer_h\right] \times \mathbb{I}_{\{t\geq 0\}} + \alpha_h + \gamma_t + \epsilon_{h,t}, \tag{3}$$

where h denotes charterer.

We present the estimation results of Eq.(3) in Table 8. Consistent with our previous results, we again find a significant drop in the usage of high-risk tankers by U.S. charterers, while the effects from U.S.-allied charterers are not statistically significant here. To put the economic magnitudes of these results in perspective, the number of tankers – later disclosed as high-risk– employed by the average U.S. charterer in July 2023 is 0.652. The estimated 0.358 drop, therefore, implies a 54.9% (0.358/0.652) reduction in usage. In the Internet Appendix F, we list the U.S. charterers to examine their usage of high-risk tankers individually. However, we don't find any clear clusters based on observable characteristics (e.g., private company vs. public company) that would suggest specific groups change their usage by more.

Finally, we examine whether selection issues affect our data: in particular, we examine whether the information shock changed the high/moderate-risk tankers' incentive to report their fixture contracts to our data vendor. In results reported in Internet Appendix F, we find no evidence to suggest any such changes in reporting by risky tankers. Note also that our ownership change data does not come from reports to Refinitiv, and is not affected by any such potential selection issues.

5 Sanctions and Disclosure in the Oil Shipping Market: A Model

In this section, we present a dynamic equilibrium model of oil shipping, and calibrate it using our key empirical results listed above. The model helps us study how the disclosure event might have differently affected sanctions violators and sanctions-complying tankers. We need the model here because we do not observe who actually violated sanctions in the data (outside of our Iranian dataset); we only observe who was labeled as high risk. Additionally,

we also use the model to understand the impact of this disclosure on the costs (and therefore profits) of sanctioned oil exporters/importers (or charterers working for them). Again, this is something that we cannot observe in the data.

5.1 The model

5.1.1 Environment

A. Players

The oil shipping sector has two groups of players: oil exporters and oil tankers. There are two types of exporters – of sanctioned oil (e.g., Iranian oil companies) and non-sanctioned oil (e.g., US oil companies), whom we denote respectively, as 'Rogue' and 'Clean'. These exporter types do not change. Note that in our analysis, we bunch together exporters and importers of sanctioned oil under the label Rogue exporters.

For oil tankers, we model the "shadow fleet" – that is, tankers that can potentially be involved in sanctions violations, but are difficult to detect. These tankers are also of two types: those that have previously violated sanctions by carrying sanctioned oil (denoted 'Bad' or B), and those that have never violated sanctions before (denoted 'Good' or G). This implies that a Bad tanker cannot turn into a Good one, and that a Good tanker will turn irreversibly into a Bad one if it engages with a Rogue exporter and carries their sanctioned oil. This assumption is motivated by the fact that tankers can be penalized for past sanctions violations. Importantly, being a Bad tanker does *not* imply that it has actually been sanctioned and put on a designated list – it only implies that this tanker has already breached sanctions in the past and *could be added* to such a list at any time.

A clean exporter may face penalties and delays if it were employing a Bad tanker at a time when the tanker is put on a sanctioned list.[11]

B. Information structure

Rogue exporters know which tankers are Bad (maybe because they have employed these tankers previously). Clean exporters, on the other hand, only know that a proportion λ of all tankers are Bad (this is common knowledge). In order to avoid engaging with Bad tankers, these exporters are supposed to do due diligence. They do so by using a detection technology which generates a binary signal, and assigns a 'Low-risk' or 'High-risk' label to a given tanker. The tankers, of course, know if they are Bad. They also know their risk label,

[11]Note that even if a sanctions clause is included in the contract with the ship owner, such an exporter can still face disruptions in its operations and additional costs, including those related to litigation (e.g., Reed Smith, 2024).

or can deduce it from the prices (freight rates) that Clean exporters quote to them. The Rogue exporters do not know these labels or the price quote that a Clean exporter might have offered a tanker. We focus on a stationary environment, where the proportion of B tankers, λ, stays constant.

C. Tanker classification

Given that there are λ Bad (B) tankers and $1-\lambda$ Good (G) tankers; the Clean exporters' detection technology classifies λ tankers as High-risk (H), and $1-\lambda$ tankers as Low-risk (L). The correlation between the true tanker type (B or G) and its risk label (H or L) is $\rho \in [0,1]$, which is a measure of the precision of the detection technology. Using this correlation, we define the following conditional probabilities

$$\theta_B \equiv \mathbb{P}[H|B] = \lambda + (1-\lambda)\rho, \quad \theta_G \equiv \mathbb{P}[H|G] = \lambda(1-\rho), \tag{4}$$

from which it also follows that

$$\mathbb{P}[B|H] = \theta_B, \quad \mathbb{P}[B|L] = \theta_G. \tag{5}$$

For tractability, we assume that this detection technology draws an H or L label for a given tanker period by period, depending only on its true type – that is, the error in detection is i.i.d. conditional on tanker type. This structure implies that persistence in the risk labels is only driven by persistence in tanker type; in particular, when a Good tanker transitions to Bad, its probability of being classified as H will increase permanently.

D. Sanction probabilities

We do not model the sanctioning authority's optimal sanctions policy, and treat the parameters related to sanction probabilities as exogenous. From the market's perspective, the probability of a tanker being sanctioned, conditional on its risk label $x \in \{L, H\}$, is

$$w_x \cdot \mathbb{I}\{\text{Previously violated sanctions}\} + \delta_x \cdot \mathbb{I}\{\text{Currently carrying sanctioned oil}\}, \tag{6}$$

where w_x (δ_x) is the sanctioning authority's screening intensity for past (current) violations. We define screening intensity as the label-dependent probability that a given tanker is examined sufficiently so that its true type is uncovered. For example, given these screening intensities, the probability that a B tanker ($\mathbb{I}\{\text{Previously violated sanctions}\} = 1$), with label H, that carries sanctioned oil in the current period ($\mathbb{I}\{\text{Currently carrying sanctioned oil}\} = 1$) will be sanctioned is $w_H + \delta_H$. This structure implies that (i) a tanker can be sanctioned for

past violations, and (ii) in addition, violating sanctions in the current period increases the probability of getting sanctioned, irrespective of past actions.

The total screening intensities, weighted by the proportion of H and L tankers, are

$$w_0 = \lambda w_H + (1-\lambda) w_L, \quad \delta_0 = \lambda \delta_H + (1-\lambda)\delta_L. \tag{7}$$

These w_0 and δ_0 depend on the resources that the sanctioning authority can spend to uncover violations, which we assume to be constant in the model. Note that the sanctioning authority's information set could be different from the market's. Therefore, w_H and w_L (as well as δ_H and δ_L) should be interpreted as the market's best guess of the authority's screening intensities based on the market's information set. In particular, different tankers with H labels may have different screening intensities under the sanctioning authority's information set, but the market cannot further distinguish between them and only expects an average intensity w_H. We assume that the two information sets are correlated, such that the market expects resources allocated by the sanctioning authority towards investigating the set of H tankers, relative to the L set, to depend on the relative proportion of bad tankers across these sets. That is:

$$w_H/w_L = \delta_H/\delta_L = \mathbb{P}[B|H]/\mathbb{P}[B|L]. \tag{8}$$

Under this assumption, if $\rho > 0$, then H labels are more likely to be associated with B tankers, i.e. $\mathbb{P}[B|H] > \mathbb{P}[B|L]$, and hence $w_H > w_L$ and $\delta_H > \delta_L$. If, on the other hand, the risk labels are pure noise, i.e., $\rho = 0$ and $\mathbb{P}[B|H] = \mathbb{P}[B|L]$, then $w_H = w_L$ and $\delta_H = \delta_L$, i.e., the H and L labels are not associated with any difference in screening intensities.

Combining equations (5), (7) and (8), we obtain

$$w_H = \frac{\theta_B}{\lambda} w_0, \quad w_L = \frac{\theta_G}{\lambda} w_0, \quad \delta_H = \frac{\theta_B}{\lambda}\delta_0, \quad \delta_L = \frac{\theta_G}{\lambda}\delta_0. \tag{9}$$

E. Timeline

We assume the following sequence of events each period:

1. Risk labels are assigned by the best existing technology to tankers by Clean exporters.

2. An aggregate mean-zero shock $\tilde{\epsilon}$ to the oil trading revenue of all exporters is realized. This shock could reflect unpredictable changes in oil prices that exporters are unable to hedge ex-ante. Besides, a tanker-specific shock $\tilde{c}_i$ for each tanker i is realized, which affects tankers' costs of doing business when dealing with Rogue exporters.

3. Clean and Rogue exporters simultaneously choose prices (i.e., freight rates) for the tankers, given their information on tanker labels/types. That is, Clean exporters choose p_L and p_H for L and H tankers, respectively, and Rogue exporters choose p_G^R and p_B^R for G and B tankers, respectively.

4. Tankers observe the prices (i.e., freight rates) given to them and choose whether to deal with Clean or Rogue exporters.

5. Each player generates a payoff and moves to the next period. Tankers live to see another period only if they are not sanctioned based on their current and past violations.

5.1.2 Exporters' optimization

The per-unit oil trade revenue for Rogue exporters, $\tilde{r}^R$, and for Clean exporters, $\tilde{r}$, are

$$\tilde{r}^R = r^R + \tilde{\epsilon}, \quad \tilde{r} = r + \tilde{\epsilon}, \tag{10}$$

where r^R and r are mean values and $\tilde{\epsilon}$ is an aggregate shock uniformly i.i.d. over $[-\sigma, \sigma]$. As equation (10) shows, we allow $\tilde{r}^R$ and $\tilde{r}$ to have different means, while they share the common shock $\tilde{\epsilon}$. Note that we only model the oil shipping market, and not the oil market itself, so we assume for simplicity a fixed ratio $\frac{r^R}{r}$. For our calibration exercise, such an assumption is justified by the fact that there were no significant changes in the discount for oil from sanctioned jurisdictions in the months after Windward's disclosure, the key event of our interest (see, e.g., Inside Shipping, 2024 – until October of 2023, that is, 3 months post-Windward's disclosure, there was no increase in the discount on Russian Ural oil).

All exporters choose prices to maximize their profits. Since there is no value to past 'relationships' between exporters and tankers in our setup, and we operate in an i.i.d. environment, exporters maximize profits period by period. Their cost consists of the price that exporters pay to tankers, as well as the expected sanctions penalty. For Rogue exporters, this penalty is z^R times the sanction probability.

Let Q_{ij}^R denote the supply of tankers of type i and risk label j that deal with Rogue exporters. Implicitly this is a function of p_G^R and p_B^R, which are optimally chosen given Clean exporters' prices. After observing the realization $\tilde{r}^R$, Rogue exporters maximize profits:

$$\begin{aligned}\max_{\{p_G^R, p_B^R\}} \; & Q_{GL}^R(\tilde{r}^R - p_G^R - \delta_L z^R) + Q_{GH}^R(\tilde{r}^R - p_G^R - \delta_H z^R) \\ & + Q_{BL}^R[\tilde{r}^R - p_B^R - (\delta_L + w_L) z^R] + Q_{BH}^R[\tilde{r}^R - p_B^R - (\delta_H + w_H) z^R].\end{aligned} \tag{11}$$

The expected sanction penalty for Clean exporters is different from that for Rogue ex-

porters. This is because (i) tankers hired by Clean exporters do not face the additional probability of being sanctioned, δ_L or δ_H, as they do not violate sanctions in the current-period, and (ii) the penalty z for a Clean exporter using a bad tanker which gets sanctioned in the current period is smaller than the z^R for Rogue exporters. For example, the entire oil cargo of a Rogue exporter could be impounded if a tanker carrying it is sanctioned; this is significantly more costly than the delays and penalties for insufficient due diligence that a Clean exporter would face.

Let Q_{ij} denote the supply of tankers of type i and risk label j that deal with Clean exporters. Implicitly, Q_{ij} is a function of p_L and p_H, which is optimally chosen given Rogue exporters' prices. After observing the realization $\tilde{r}$, Clean exporters maximize:

$$\max_{\{p_L, p_H\}} \; Q_{GL}(\tilde{r} - p_L) + Q_{GH}(\tilde{r} - p_H) + Q_{BL}(\tilde{r} - p_L - w_L z) + Q_{BH}(\tilde{r} - p_H - w_H z). \quad (12)$$

The key mechanism in the model that makes detection accuracy important is that practical limits to enforcement limit the size of z^R; for example, if the sanctioning authority seizes a tanker carrying sanctioned oil, they can at most impound the tanker and its cargo. If z^R could be increased without such bounds, then violations could be prevented even if detection was noisy; in this sense our model is similar in spirit to Becker's classic intuition on crime and punishment (Becker, 1968).

5.1.3 Tankers' optimization

Tankers are price-takers – they observe the prices (freight rates) quoted to them by Clean and Rogue exporters and then decide which of these exporters to deal with. Because a G tanker will turn irreversibly into a B tanker if it engages with Rogue exporters, tankers need to consider their continuation values when making decisions.

Let V_G and V_B denote the unconditional values of G and B tankers, and $V_{GL}, V_{GH}, V_{BL}, V_{BH}$

denote the values conditional on risk labels. The Bellman equations for tankers are

$$
\begin{aligned}
V_{GL} &= \mathbb{E}\left[\max\begin{Bmatrix} p_L + \beta V_G \\ p_G^R - \tilde{c}_i - \delta_L s + \beta V_B \end{Bmatrix}\right], \\
V_{GH} &= \mathbb{E}\left[\max\begin{Bmatrix} p_H + \beta V_G \\ p_G^R - \tilde{c}_i - \delta_H s + \beta V_B \end{Bmatrix}\right], \\
V_{BL} &= \mathbb{E}\left[\max\begin{Bmatrix} p_L - w_L s + \beta V_B \\ p_B^R - \tilde{c}_i - (w_L + \delta_L) s + \beta V_B \end{Bmatrix}\right], \\
V_{BH} &= \mathbb{E}\left[\max\begin{Bmatrix} p_H - w_H s + \beta V_B \\ p_B^R - \tilde{c}_i - (w_H + \delta_H) s + \beta V_B \end{Bmatrix}\right],
\end{aligned}
$$

where V_G and V_B are given by

$$
V_G = \theta_G V_{GH} + (1 - \theta_G) V_{GL}, \quad V_B = \theta_B V_{BH} + (1 - \theta_B) V_{BL}. \tag{13}
$$

When dealing with Rogue exporters, tanker i incurs in each period a random cost $\tilde{c}_i$ which we assume to be uniformly i.i.d. over $[0, \bar{c}]$. This cost is only observed by the tanker itself before it decides which exporter to deal with. It allows us to obtain interior solutions for tanker supply. Such a cost may stem from efforts to circumvent detection – which can depend on the interaction between tanker location, route, and changes in evasion technology, or from uncertainties concerning financing or other relations with third parties who seek to avoid falling under secondary sanctions.

The penalty imposed on a Bad tanker when sanctioned is s. Here we assume for simplicity that it will lose its entire value if sanctioned. Given discounting, this loss of value means that $s = \beta V_B$.

Critical values of $\tilde{c}_i$ at which tanker i would be indifferent between dealing with a Clean or Rogue exporter are as below.

$$
\begin{aligned}
c_{GL} &= p_G^R - p_L - \delta_L s + \beta(V_B - V_G), \\
c_{GH} &= p_G^R - p_H - \delta_H s + \beta(V_B - V_G), \\
c_{BL} &= p_B^R - p_L - \delta_L s, \\
c_{BH} &= p_B^R - p_H - \delta_H s.
\end{aligned}
$$

For $\tilde{c}_i$ below (above) such a critical value the tanker prefers to deal with a Rogue (Clean)

exporter. Using these critical values, we derive that

$$V_G = \frac{\theta_G \mathbb{E}\left[\frac{(c_{GH})^2}{2\bar{c}} + p_H\right] + (1-\theta_G)\mathbb{E}\left[\frac{(c_{GL})^2}{2\bar{c}} + p_L\right]}{1-\beta}, \tag{14}$$

$$V_B = \frac{\theta_B \mathbb{E}\left[\frac{(c_{BH})^2}{2\bar{c}} + p_H\right] + (1-\theta_B)\mathbb{E}\left[\frac{(c_{BL})^2}{2\bar{c}} + p_L\right] - [\theta_B w_H + (1-\theta_B) w_L]s}{1-\beta}, \tag{15}$$

where the expectation is over next period's aggregate shock $\tilde{\epsilon}$.

5.1.4 Equilibrium

A stationary equilibrium is characterized by: (I) tankers' optimal decision rule based on critical values $c_{GL}(\tilde{\epsilon})$, $c_{GH}(\tilde{\epsilon})$, $c_{BL}(\tilde{\epsilon})$, $c_{BH}(\tilde{\epsilon})$ where $\tilde{\epsilon}$ refers to the aggregate shock in the current period, and, (II) Rogue and Clean exporters' optimal prices $p_G^R(\tilde{\epsilon})$, $p_B^R(\tilde{\epsilon})$, $p_L(\tilde{\epsilon})$, $p_H(\tilde{\epsilon})$, such that: (i) tankers maximize their values, and exporters maximize their profits, (ii) exporters' beliefs about tanker types or risk labels are consistent with tankers' decisions in aggregate, and (iii) markets clear, as below:
The market clearing conditions for the Rogue exporters are

$$\begin{aligned}
Q_{GL}^R &= (1-\lambda)(1-\theta_G)\left(\frac{c_{GL}}{\bar{c}}\right), \\
Q_{GH}^R &= (1-\lambda)\theta_G\left(\frac{c_{GH}}{\bar{c}}\right), \\
Q_{BL}^R &= \lambda(1-\theta_B)\left(\frac{c_{BL}}{\bar{c}}\right), \\
Q_{BH}^R &= \lambda\theta_B\left(\frac{c_{BH}}{\bar{c}}\right)
\end{aligned}$$

and for the Clean exporters they are

$$\begin{aligned}
Q_{GL} &= (1-\lambda)(1-\theta_G)\left(1-\frac{c_{GL}}{\bar{c}}\right), \\
Q_{GH} &= (1-\lambda)\theta_G\left(1-\frac{c_{GH}}{\bar{c}}\right), \\
Q_{BL} &= \lambda(1-\theta_B)\left(1-\frac{c_{BL}}{\bar{c}}\right), \\
Q_{BH} &= \lambda\theta_B\left(1-\frac{c_{BH}}{\bar{c}}\right).
\end{aligned}$$

5.2 Model calibration

The average number of trips in one year in our fixtures sample is 3.4. We set the discount factor β accordingly to 0.9, which incorporates both the time discounting and depreciation of

tanker values.[12] Based on (CRS, 2024), we set the proportion λ of Bad tankers to 0.213. We normalize the Clean exporters' mean oil trade revenue (profits before shipping costs, to be more precise) r per voyage to one. This revenue amounts to 15.5 million for an average-size oil tanker (i.e., 0.8 million barrels), at an oil price of \$70 per barrel, and with a 27.7% before-shipping profit margin (which was the operating margin in the oil industry for 2023 Q2, as per CSI Market, 2023). For the Rogue exporters we assume a four-dollar per barrel discount (as observed for Russian exports to India in the second half of 2023, as per Reuters, Sept. 2023), and hence their mean oil trade revenue is $r^R = (70 \times 0.277 - 4)/(70 \times 0.277) = 0.794$.

To calibrate the sanction penalty z for Clean exporters when dealing with Bad tankers, we note that OFAC's civil monetary penalties are typically determined in relation to the amount of the transaction underlying the sanctions violation (CFR, Appendix A to Part 501). Thus we set z equal to the freight cost they pay, i.e., around \$4 million in 2023 (e.g., Inside Shipping, 2024), which implies $z = 4/15.5 = 0.258$. We set the sanction penalty z^R for Rogue exporters based on the median value from actual cases of seized oil (e.g., US Attorney's Office, 2024), which gives $z^R = 50/15.5 = 3.226$. We set the classification precision with which Clean exporters can detect sanctions-violating tankers by converting the pseudo-R^2 of sanction predictability in Figure 8 to the correlation coefficient ρ in the model. The pseudo-R^2 based on public data is 7.5%, which implies a correlation ρ^{pre} of 0.296, while the pseudo-R^2 after adding Windward's risk labels to the information set is 17.2%, implying a correlation ρ^{post} of 0.453 (in what follows superscripts *pre* and *post* denote the pre- and post-disclosure periods).

We calibrate the screening intensities (w_0 and δ_0), revenue volatility (σ), and maximum cost ($\bar{c}$) by matching six moments. These moments are (i) the proportion of sanctioned tankers, (ii) the annualized volatilities of low-risk and (iii) high-risk tankers' fixtures, relative to their mean, (iv) the annualized volatility of the top US exporter/charterers' revenue relative to its mean (proxy for that of Clean exporters), (v) the disclosure's effect on high-risk tankers' fixtures (from the DiD estimation in Table 5), and (vi) the disclosure's effect on high-risk tankers' routes (from the DiD estimation in Table 6). We construct model analogs of these moments as follows:

First, the number of sanctioned tankers in the model is $Q_{Sanc} = Q_{BL}w_L + Q_{BH}w_H + Q^R_{GL}\delta_L + Q^R_{GH}\delta_H + Q^R_{BL}(w_L + \delta_L) + Q^R_{BH}(w_H + \delta_H)$ (since we normalize the total number of tankers to one, this will give us the proportion of sanctioned tankers). The volatilities of low-risk and high-risk tankers' fixtures are $\text{std}(p_L)$ and $\text{std}(p_H)$, respectively.

Second, Clean exporters' total revenue in the model equals $(Q_{GL} + Q_{GH} + Q_{BL} + Q_{BH})\tilde{r}$.

[12] For example, assuming a time-discounting rate of 0.95, a scrap value of 10%, and a 15-year life cycle gives $0.95 \times \exp(\log(0.1)/(15 \times 3.4)) \approx 0.9$.

We get its empirical equivalent by aggregating the revenues of the largest US oil companies from their 10-K reports. We use the top 12 publicly traded US oil companies. The full list of large US companies involved in oil shipping, including those that we use and also some private ones, is in the Internet Appendix Table A-4.

Third, the model analog for the disclosure effect on high-risk tankers' fixtures is

$$\frac{\mathbb{E}[p^{post} - p^{pre}|\mathrm{H}^{post}]}{\mathbb{E}[p^{pre}|\mathrm{H}^{post}]}, \tag{16}$$

where p is fixture and we use superscripts "pre" or "post" to indicate periods – for example, H^{post} is a tanker that is labeled as high-risk in the post-period. The average pre-period fixture, conditional on a post-period high-risk label is

$$\mathbb{E}[p^{pre}|\mathrm{H}^{post}] = \mathbb{P}[\mathrm{L}^{pre}|\mathrm{H}^{post}] \times \mathbb{E}[p_L^{pre}] + \mathbb{P}[\mathrm{H}^{pre}|\mathrm{H}^{post}] \times \mathbb{E}[p_H^{pre}], \tag{17}$$

where $\mathbb{E}[p_L^{pre}]$ and $\mathbb{E}[p_H^{pre}]$ are price averages taken over the pre-period aggregate shock $\tilde{\epsilon}$ and $\mathbb{P}[.|.]$ is conditional probability. Noting that $\mathbb{P}[\mathrm{L}^{pre}|\mathrm{H}^{post}] = 1 - \mathbb{P}[\mathrm{H}^{pre}|\mathrm{H}^{post}]$, we calculate $\mathbb{P}[\mathrm{H}^{pre}|\mathrm{H}^{post}]$ by splitting it into three cases based on tanker types: (i) G^{pre} and G^{post}, (ii) G^{pre} and B^{post}, and (iii) B^{pre} and B^{post}. The formula is shown in the Internet Appendix.

Finally, we use the probability of sending AIS signals from areas near sanctioned countries (i.e., the estimation in Table 6) as a proxy for the probability of dealing with Rogue exporters in the model. We define b_H and b_L as the probabilities for H and L tankers of dealing with Rogue exporters:

$$\begin{aligned} b_H &= \mathbb{P}[\mathrm{B}|\mathrm{H}] \times \mathbb{P}[\tilde{c}_i < c_{BH}] + \mathbb{P}[\mathrm{G}|\mathrm{H}] \times \mathbb{P}[\tilde{c}_i < c_{GH}] = \theta_B \frac{c_{BH}}{\bar{c}} + (1-\theta_B)\frac{c_{GH}}{\bar{c}}, \\ b_L &= \mathbb{P}[\mathrm{B}|\mathrm{L}] \times \mathbb{P}[\tilde{c}_i < c_{BL}] + \mathbb{P}[\mathrm{G}|\mathrm{L}] \times \mathbb{P}[\tilde{c}_i < c_{GL}] = \theta_G \frac{c_{BL}}{\bar{c}} + (1-\theta_G)\frac{c_{GL}}{\bar{c}}. \end{aligned}$$

Based on these definitions, the model analog for the disclosure effect on high-risk tankers' routes is

$$\frac{\mathbb{E}[b^{post} - b^{pre}|\mathrm{H}^{post}]}{\mathbb{E}[b^{pre}|\mathrm{H}^{post}]} \tag{18}$$

where

$$\mathbb{E}[b^{pre}|\mathrm{H}^{post}] = \mathbb{P}[\mathrm{L}^{pre}|\mathrm{H}^{post}] \times \mathbb{E}[b_L^{pre}] + \mathbb{P}[\mathrm{H}^{pre}|\mathrm{H}^{post}] \times \mathbb{E}[b_H^{pre}], \tag{19}$$

$\mathbb{E}[b_L^{pre}]$ and $\mathbb{E}[b_H^{pre}]$ are the averages of the respective probabilities taken over the pre-period aggregate shock $\tilde{\epsilon}$. The conditional probability $\Pr[\mathrm{H}^{pre}|\mathrm{H}^{post}]$ is now calculated as specified above.

We solve the model by discretizing the aggregate shock $\tilde{\epsilon}$ on a grid. The key parameter

is the classification precision, ρ, which captures the information shock from Windward's disclosure.

Panel B in Table 9 shows that the model can fit the data quite well. The data values vs. model analogs for the proportion of sanctioned tankers, the annualized volatilities of low-risk and high-risk tankers' fixtures, and the annualized volatility of top US charterers' revenue are 0.5% vs. 0.5%, 0.77 vs. 0.77, 0.81 vs. 0.83, and 0.30 vs. 0.21, respectively. More importantly, the model can reproduce the disclosure effect on high-risk tankers' fixture changes, i.e., a 13% drop in the data vs. 11% in the model, and route changes, i.e., a 17% drop in the data vs. 20% in the model.

Importantly, the variable values in the model can match what we know about their real-world values from the media, financial market data, or other sources. First, the volatility of oil trade revenue, $\sigma = 0.365$, derived from moments matching, is consistent with the data – the Oil Volatility Index OVX, which measures the oil price volatility, and hence, earnings volatility, is around 0.40 over 2022-2023. Second, a key model implication from Panel D in Table 9 is that the Rogue exporters pay a freight rate premium of around 110%-125% relative to the Clean exporters (i.e., $(\mathbb{E}[p_B^R] + \mathbb{E}[p_G^R])/(\mathbb{E}[p_H] + \mathbb{E}[p_L]) - 1$). This model value matches closely the sanctions premium reported in the media for Russian oil delivered in 2023 to Indian and Chinese ports (e.g., Inside Shipping, 2024, which reports calculations from Argus Media). Finally, the model-based tanker values between \$37 and \$44 million reported in Table 9 Panel D also match their actual values – in 2023, a 15-year-old tanker had an average value of \$45 million in the second-hand market (using values for VLCC, Suezmax and Aframax tankers as per Xclusiv, 2023), and a 20-year old tanker had a value of \$30 million.

Overall, the closeness of these parameter values to other estimates of similar quantities in the media provides assurance on taking quantities generated within the scope of this model seriously.

5.3 Model implications

Figure 12 illustrates several implications of the increase in classification precision derived from our model. In each subplot, the two vertical dashed lines indicate the pre- and post-Windward precisions, as measured by pseudo-R^2, which are 7.5% and 17.2%, respectively.

We first study implications for Rogue exporters, which we can only observe in the model. Panel A of Figure 12 shows the relation between classification precision, and the equilibrium freight prices for Rogue exporters (i.e., the fixtures $\mathbb{E}[p_G^R]$ and $\mathbb{E}[p_B^R]$, with expectations over the current-period aggregate shock). First, the Rogue exporters' price for hiring G (B)

tankers increases (decreases) when classification becomes more precise. This is because the screening intensity of past activity for B tankers is now higher on average (as B tankers are more likely to be labeled as H, and w_H is now higher), which makes them less attractive. At the same time, G tankers do not have past violations and, therefore, become even more attractive.

Panel B shows the relation between classification precision and tankers' probabilities of dealing with Rogue exporters. First, for B tankers this probability is higher than for G tankers. Second, within each type G and B, H tankers are less likely to deal with Rogue exporters than L tankers. This is because H tankers are under higher screening intensity, i.e., $w_H > w_L$ and $\delta_H > \delta_L$ when $\rho > 0$, which implies a higher expected cost of violating sanctions for these tankers. Finally, as classification precision increases, the probability of dealing with Rogue exporters decreases more for H tankers relative to L tankers. This is because the screening intensity increases more for H tankers, as per Equation 8 (since $\mathbb{P}[B|H]$ increases, and $\mathbb{P}[B|L]$ decreases with classification precision).

The next two panels of Figure 12 shed light on aggregate implications for Rogue exporters. Panel C shows that the per-tanker spending of Rogue exporters increases, which is due mostly to the increase in freight rates for G tankers paid by Rogue exporters (from Panel A in the figure). This increase is not offset by the drop in freight rates that they pay to B tankers because, overall, Rogue exporters employ more G tankers – these are in total about four times more than the B tankers (λ is 0.213), while B tankers are at most twice as likely to be employed by the Rogue exporters (from Panel B).

Panel D shows that with higher classification precision, the total number of oil tankers engaging with Rogue exporters decreases. This follows from Panel B. Quantitatively, the information shock accounts for an 11.7% drop in Rogue exporters' total hiring of tankers, when pseudo-R^2 increases from the pre-Windward to post-Windward precision. This shortfall, of course, can translate into a drop in exports of sanctioned oil itself, but the magnitude of that drop may not be one-for-one. This is because some of these tankers can be replaced by other means, e.g., through oil pipelines – which will now have to overwork relative to the pre-period equilibrium, potentially leading to higher costs of transport and higher depreciation of these alternative transportation assets.

Next, Panel E of Figure 12 shows the relation between classification precision and the equilibrium freight prices (i.e., the fixtures $\mathbb{E}[p_L]$ and $\mathbb{E}[p_H]$) for Clean exporters. Both these prices decrease in response to better classification, because now fewer tankers overall deal with Rogue exporters, increasing tanker supply for Clean exporters. At the same time, $\mathbb{E}[p_H]$ decreases more than $\mathbb{E}[p_L]$. This is because as precision increases, the H label indicates a higher probability of a B tanker, so Clean exporters require a larger discount for hiring such a

tanker. The left-most point of the panel takes this intuition to the extreme: when pseudo-R^2 is zero and classification is pure noise, Clean exporters set $\mathbb{E}[p_H]$ and $\mathbb{E}[p_L]$ to be equal.

Panel F shows how expected freight prices from Clean exporters change with classification precision. These prices for G and B tankers are given by

$$\mathbb{E}[p|G] = \theta_G \mathbb{E}[p_H] + (1 - \theta_G)\mathbb{E}[p_L], \quad \mathbb{E}[p|B] = \theta_B \mathbb{E}[p_H] + (1 - \theta_B)\mathbb{E}[p_L].$$

As classification becomes more precise, the decrease in θ_G and increase in θ_B (see Equation 5) makes G (B) tankers' expected price closer to $\mathbb{E}[p_L]$ ($\mathbb{E}[p_H]$).

How to explain the counter-intuitive result that the increase in classification precision diminishes the freight rates received also by G tankers, as shown in Panel F? We note, from Panel B, that with higher precision Rogue exporters increasingly avoid B tankers, fearing a higher chance of detection – their usage of BH tankers drops while BL increases slightly; however, there are more BH tankers than BL (as long as $\rho > 0$). Therefore, more B tankers now seek employment with Clean exporters, which boosts the supply available for the latter and lowers the earnings for G tankers who work with them. This drop is only partly offset by Rogue exporters' willingness to now pay more for these G tankers (Panel A), because the Rogue market is considerably smaller and continues to shrink (Panel D). Tanker values show a correspondingly declining pattern (from Table 9, Panel D), because they reflect discounted values of future expected prices.

Finally, in Table 10, we estimate the aggregate (annual) effect of the information shock on the four relevant parties in terms of dollar values. We focus on price effects and fix quantities (i.e., the Qs in the model) to pre-Windward levels. Alternatively, we could have assumed that Rogue exports drop one-for-one with the fall in their usage of oil tankers, but we think such an assumption might exaggerate effects if Rogue exporters ship the same oil using alternative means, e.g., pipelines. Of course, one could estimate the effect of such changes in transporting methods on their cost of exports if one knew how much these alternative means cost – but we do not have this information. So we estimate costs and benefits by fixing overall export quantities, and let the information disclosure affect the market only through prices.

We analyze the changes in costs/benefits for exporters and tankers between the pre- and post-Windward periods. We find that the shipping cost for Rogue exporters increases by 2% (from 9.0 to 9.2 mln per tanker-trip) due to the increase in p_G^R, while the cost for Clean exporters decreases by 6.3% (from 4.2 to 4.0 mln per tanker-trip) due to the decreases in both p_L and p_H. On the tanker side, the fixture income decreases by 5.7% (from 4.9 to 4.7 mln per tanker-trip) and 3.3% (from 5.0 to 4.9 mln per tanker-trip) for B and G

tankers, respectively. To estimate the aggregate annual effect, we do the following back-of-the-envelope calculation: we multiply the above dollar values by the number of tankers and 3.4 trips per year (as explained earlier). Therefore, Windward's disclosure leads to $780 million and $1,536 million annual losses for Rogue exporters and B tankers, respectively. These losses accrue as gains for the sanctions-compliant agents, with a $2.3 billion overall gain for Clean exporters and G tankers combined.

Taken together, these results suggest that Rogue exporters and Bad tankers are indeed adversely impacted by Windward's disclosure. As this is a key goal for sanctions enforcement, our exercise here makes a clear case for disclosing more precise information on tanker sanction risks to the market.

6 Implications and Conclusion

Our results indicate that Windward's sudden disclosure of information identifying suspected violators negatively impacted both rogue exporters and sanctions-violating tankers. Whether these findings suggest a need to reform sanctions-related information policy depends, of course, on whether our estimates point to a market failure. After all, Windward successfully aggregated various types of information to more accurately identify suspects, and eventually made this information public. Does this imply that the market was able to solve its own problems with identifying sanctions violators, meaning no policy changes are necessary?

One way to approach this question is by comparing the value of Windward's information to the market versus its value to sanctioning authorities. On the one hand, Windward's total market capitalization was $47 million just before its list of suspected sanctions-violating tankers was made public. This $47 million reflects the discounted value of Windward's future profits, which in turn represents the market's willingness to pay for its information. On the other hand, following Windward's disclosure, sanctioned exporters faced an increase in shipping costs of approximately $780 million per year, and sanction-violating tankers experienced a reduction in freight rates exceeding $1.5 billion (from Section 5).

The key question then becomes: how much should a sanctioning authority that is focused on enforcement be willing to pay to impose a $2.3 billion annual cost on violators? If this amount is more than the $47 million the market was willing to pay, then our findings indeed suggest a need for change in information provision.

References

Abadie, A. (2005), 'Semiparametric Difference-in-Differences Estimators', *Review of Economic Studies* **72**(1), 1–19.

Aumann, R. J. (1976), 'Agreeing to disagree', *Annals of Statistics* **4**(6), 1236–1239.

Babina, T., Hilgenstock, B., Itskhoki, O., Mironov, M. and Ribakova, E. (2023), 'Assessing the impact of international sanctions on Russian oil exports', *Available at SSRN* .

Becker, G. (1968), 'Crime and punishment: An economic approach', *Journal of Political Economy* **76**, 169–217.

Brancaccio, G., Kalouptsidi, M. and Papageorgiou, T. (2020), 'Geography, transportation, and endogenous trade costs', *Econometrica* **88**(2), 657–691.

Brancaccio, G., Kalouptsidi, M. and Papageorgiou, T. (2023), 'The impact of oil prices on world trade', *Review of International Economics* **31**(2), 444–463.

Brown, P. (2020), *Oil market effects from US economic sanctions: Iran, Russia, Venezuela*, Vol. 3, Congressional Research Service Washington, DC, USA.

Clayton, C., Maggiori, M. and Schreger, J. (2023), A framework for geoeconomics, Working Paper 31852, National Bureau of Economic Research.

Crozet, M., Hinz, J., Stammann, A. and Wanner, J. (2021), 'Worth the pain? Firms' exporting behaviour to countries under sanctions', *European Economic Review* **134**.

Dellavigna, S. and La Ferrara, E. (2010), 'Detecting illegal arms trade', *American Economic Journal: Economic Policy* **2**(4), 26–57.

Dimmock, S. G. and Gerken, W. C. (2012), 'Predicting fraud by investment managers', *Journal of Financial Economics* **105**(1), 153–173.

Dizaji, S. F. and van Bergeijk, P. A. G. (2013), 'Potential early phase success and ultimate failure of economic sanctions: A VAR approach with an application to Iran', *Journal of Peace Research* **50**, 721–736.

Duggan, M. and Levitt, S. D. (2002), 'Winning isn't everything: Corruption in sumo wrestling', *American Economic Review* **92**(5), 1594–1605.

Early, B. R. and Preble, K. (2019), 'Enforcing Economic Sanctions: Analyzing How OFAC Punishes Violators of U.S. Sanctions', *Working Paper* .

Eaton, J. and Engers, M. (1992), 'Sanctions', *Journal of Political Economy* **100**, 899–928.

Eaton, J. and Engers, M. (1999), 'Sanctions: Some simple analytics', *American Economic Review* **89**, 409–414.

Efing, M., Goldbach, S. and Nitsch, V. (2023), 'Freeze! Financial sanctions and bank responses', *Review of Financial Studies* **36**, 4417–4459.

Farzanegan, M. R. (2011), 'Oil revenue shocks and government spending behavior in Iran', *Energy Economics* **33**(6), 1055–1069.

Fisman, R. J., Marcolongo, G. and Wu, M. (2024), 'The undoing of economic sanctions: Evidence from the Russia-Ukraine conflict', *Working Paper* .

Fisman, R. and Wang, Y. (2017), 'The Distortionary Effects of Incentives in Government: Evidence from China's "Death Ceiling" Program', *American Economic Journal: Applied Economics* **9**, 202–218.

Fisman, R. and Wei, S.-J. (2004), 'Tax rates and tax evasion: evidence from "missing imports" in China', *Journal of Political Economy* **112**(2), 471–496.

Fisman, R. and Wei, S.-J. (2009), 'The smuggling of art, and the art of smuggling: Uncovering the illicit trade in cultural property and antiques', *American Economic Journal: Applied Economics* **1**(3), 82–96.

Gordon, J. (2016), 'Extraterritoriality: Issues of Overbreadth and the Chilling Effect in the Cases of Cuba and Iran', *Harvard International Law Journal* **57**.

Griffin, J. and Kruger, S. (2023), 'What is Forensic Finance?', *Available at SSRN* .

Griffin, J. M. and Maturana, G. (2016), 'Who facilitated misreporting in securitized loans?', *The Review of Financial Studies* **29**(2), 384–419.

Griffin, J. M. and Shams, A. (2018), 'Manipulation in the VIX?', *Review of Financial Studies* **31**(4), 1377–1417.

Haidar, J. I. (2017), 'Sanctions and export deflection: evidence from Iran', *Economic Policy* **32**, 319–355.

Hamilton, J. (1983), 'Oil and the macroeconomy since World War II', *Journal of Political Economy* **38**, 228–248.

Heckman, J. J., Ichimura, H. and Todd, P. E. (1997), 'Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme', *Review of Economic Studies* **64**(4), 605–654.

Hong, H. and Stein, J. (1999), 'A unified theory of underreaction, momentum trading, and overreaction in asset markets', *Journal of Finance* **54**, 2143–2184.

Hong, H., Torous, W. and Valkanov, R. (2007), 'Do industries lead stock markets?', *Journal of Financial Economics* **83**, 367–396.

Hong, H. and Yogo, M. (2012), 'What does futures market interest tell us about the macroeconomy and asset prices?', *Journal of Financial Economics* **105**, 473–490.

Hsieh, C.-T. and Moretti, E. (2006), 'Did Iraq cheat the United Nations? Underpricing, bribes, and the oil for food program', *Quarterly Journal of Economics* **121**, 1211–1248.

Huynh, L. D. T., Hoang, K. and Ongena, S. (2023), 'The impact of foreign sanctions on firm performance in Russia', *Swiss Finance Institute Research Paper* (23-115).

Itskhoki, O. and Mukhin, D. (2022), Sanctions and the exchange rate, Technical report, National Bureau of Economic Research.

Jacob, B. A. and Levitt, S. D. (2003), 'Rotten apples: An investigation of the prevalence and predictors of teacher cheating', *Quarterly Journal of Economics* **118**(3), 843–877.

Kaplow, L. (1990), 'Optimal Deterrence, Uninformed Individuals, and Acquiring Information about Whether Acts Are Subject to Sanctions', *Journal of Law, Economics, and Organization* **6**(1), 93–128.

Kilpatrick, R. (2022), 'Monitoring sanctions compliance at sea', *Northwestern Journal of International Law and Business* **42**, 221–251.

Lacy, D. and Niou, E. M. S. (2004), 'A theory of economic sanctions and issue linkage: The roles of preferences, information, and threats', *Journal of Politics* **66**(1), 25–42.

Lastauskas, P., Proškutė, A. and Žaldokas, A. (2023), 'How do firms adjust when trade stops?', *Journal of Economic Behavior & Organization* **216**, 287–307.

Liberti, J. M., Seru, A. and Vig, V. (2016), ''Information, Credit, and Organization''.

Mamonov, M. and Pestova, A. (2023), 'The price of war: Macroeconomic and cross-sectional effects of sanctions on Russia'.

Marion, J. and Muehlegger, E. (2008), 'Measuring illegal activity and the effects of regulatory innovation: Tax evasion and the dyeing of untaxed diesel', *Journal of Political Economy* **116**(4), 633–666.

Mulder, N. (2022), 'The Sanctions Weapon', *Finance & Development June 2022* .

Peress, J. (2014), 'The media and the diffusion of information in financial markets: Evidence from newspaper strikes', *Journal of Finance* **69**(5), 2007–2043.

Ready, R. C. (2018), 'Oil consumption, economic growth, and oil futures: The impact of long-run oil supply uncertainty on asset prices', *Journal of Monetary Economics* **94**, 1–26.

Rosenbaum, P. R. and Rubin, D. B. (1983), 'The central role of the propensity score in observational studies for causal effects', *Biometrika* **70**(1), 41–55.

Siddiquee, S. H. M. and van Bergeijk, P. (2012), Reconsidering Economic Sanctions Reconsidered: Detailed Analysis of the Peterson Institute Sanction Database, ISS Working paper, Erasmus University.

Weng, J., Li, G. and Zhao, Y. (2022), 'Detection of abnormal ship trajectory based on the complex polygon', *Journal of Navigation* **75**(4), 966–983.

Wolsing, K., Roepert, L., Bauer, J. and Wehrle, K. (2022), 'Anomaly Detection in Maritime AIS Tracks: A Review of Recent Approaches', *Journal of Marine Science and Engineering* **10**, 1–19.

Table 1: Summary statistics of detection samples

This table reports the summary statistics for the detection samples in Section 3. We separately detect spoofing and dark activity. Panel A reports the sample sizes and panel B reports the mean and standard deviation of predictors. The construction of the predictors is described in Section 3.3. "-" in panel B indicates that the sample does not include the respective predictor.

Panel A: Sample Size		
	Spoofing Detection	Dark Detection
No. observations	17,496 (tanker-day)	1,552 (dark periods)
No. unique tankers	1,481	746
No. unique violators	17	20

Panel B: Predictors				
	Spoofing Detection		Dark Detection	
	Mean	Std	Mean	Std
Identity Change	0.118	0.322	0.144	0.352
Risky Flag	0.263	0.440	0.288	0.453
Ship-to-ship Transfer				
• anchor 10 km away from ports	0.218	0.413	0.363	0.481
• anchor 20 km away from ports	0.081	0.273	0.137	0.344
• draft change 10 km away from ports	0.036	0.186	0.151	0.358
• draft change 20 km away from ports	0.027	0.161	0.119	0.324
Irregular Trajectory				
• distance cutoff = 1 km	0.050	0.218	0.053	0.225
• distance cutoff = 2 km	0.050	0.218	0.068	0.251
• distance cutoff = 5 km	0.050	0.218	0.079	0.269
DBSCAN Outlier				
• distance parameter = 10 km	0.131	0.338	0.318	0.466
• distance parameter = 20 km	0.049	0.216	0.142	0.349
Dark Activity				
• at least 2-day dark period	-	-	0.406	0.491
• at least 3-day dark period	-	-	0.244	0.429
Satellite Images				
• 1-day window	0.010	0.101	-	-
• 3-day window	0.028	0.165	-	-
• 5-day window	0.042	0.200	-	-
• during dark period	-	-	0.025	0.157

Table 2: Performance of ML detection models

The hyper-parameters of the models are optimally determined by 10-fold cross-validation and reported in Internet Appendix Table A-1. Panel A (Panel B) reports the model performance using only AIS data (both AIS and satellite data). We separately report the detection of spoofing and dark activity, and "total" reports the simple average of these two. The implied false positive rate is calculated from the pseudo-R^2, with $\bar{p} = 2.23\%$ as estimated from our data and a risk level $1 - \beta = 5\%$ (as per Internet Appendix B). Panel C illustrates the importance of individual categories of predictors. Δpseudo-R^2 for a category is the difference between the pseudo-R^2 of a full model, as those reported in panel B, and the same model that sets this particular category to zero (note that all our predictors are zero-one indicators). When calculating Δpseudo-R^2, we treat negative pseudo-R^2 or negative Δpseudo-R^2 as zero. We report the simple averages of the respective Δpseudo-R^2's for the tree and neural network models, as well as spoofing and dark detection. The last column of this panel shows the average increase in the implied false positive rate when setting the respective category to zero.

Panel A: Model Performance Using Only AIS Data			
Detect	Model	Pseudo-R^2	Implied false positive rate
Total	Tree	**9.1%**	**52.9%**
	NN	**10.2%**	**49.8%**
Spoofing	Tree	6.7%	60.5%
	NN	7.2%	58.9%
Dark	Tree	11.5%	46.3%
	NN	13.3%	42.2%
Panel B: Model Performance Using AIS and Satellite Data			
Detect	Model	Pseudo-R^2	Implied false positive rate
Total	Tree	**15.2%**	**38.0%**
	NN	**16.3%**	**35.8%**
Spoofing	Tree	11.8%	45.7%
	NN	13.8%	40.9%
Dark	Tree	18.7%	31.6%
	NN	18.8%	31.4%
Panel C: Predictor Importance			
Category		ΔPseudo-R^2	ΔImplied false positive rate
Satellite Detection		6.2%	14.7%
Identity Change		3.7%	8.0%
Risky Flag		2.8%	6.1%
Irregular Trajectory		0.2%	0.3%
Ship-to-ship Transfer		6.2%	14.8%
DBSCAN Outlier		5.3%	12.2%
Dark Activity		9.4%	24.5%

Table 3: Comparison between Windward's disclosure and ML-model detection

The first two lines in this table show the total number of tankers that violated Iranian sanctions during 2021 January, as per our proprietary dataset, and how many among them are reported on Windward's list as high/moderate risk. The bottom panel shows the performance of our decision tree and neural network models. We use either only AIS-based predictors or combine them with predictors based on satellite data. The set of predictors is as described in Section 3. The optimal hyper-parameters of the models, e.g., tree depth or number of layers of a neural network, are determined via 10-fold cross-validation. We show separately the number of sanctions-violating tankers that engage in spoofing and going dark that our models detect; we also report the total number of detections for both types of behavior – the totals can be smaller than the sum of the two types of behavior because a tanker can engage in both. Results are shown for different confidence levels ($cl = 99\%$, 95%, and 90%) – for example, under a 95% confidence level no more than 5% of complying tankers can be mislabelled as violators.

Total Violators:	33	
on Windward's list:	**27**	

			Number of violators detected		
Detect	Model	Predictors	$cl = 99\%$	$cl = 95\%$	$cl = 90\%$
Total	Tree	Satellite + AIS	**5**	**12**	**17**
	NN	Satellite + AIS	**5**	**13**	**15**
	Tree	Only AIS	3	8	14
	NN	Only AIS	0	5	13
Spoofing	Tree	Satellite + AIS	3	5	6
	NN	Satellite + AIS	4	5	6
	Tree	Only AIS	1	2	5
	NN	Only AIS	0	3	4
Dark	Tree	Satellite + AIS	3	9	14
	NN	Satellite + AIS	1	10	12
	Tree	Only AIS	2	6	10
	NN	Only AIS	0	4	11

Table 4: Summary statistics of Windward's tankers sample

This table reports the summary statistics for the sample of tankers we use in Section 4 to examine the effects of the disclosure event. Panel A reports sample sizes and tanker classification. Panel B reports the balance tests on tanker characteristics between treated (i.e., high- and moderate-risk) and control (i.e., low-risk) groups. To construct the control group, we conduct propensity score matching (PSM) for each treated tanker within its tanker type, i.e., we require matched tanker(s) to have the same tanker type. The propensity score is calculated by regressing the high/moderate-risk tanker indicator on pre-period tanker characteristics using ML methods: average outputs from a decision tree and a neural network. The variables used to construct the propensity score are the same as the list in panel B. We calculate matching weights based on a Gaussian kernel with a bandwidth of 0.05. The variable differences and corresponding t-value are reported. For this tanker-level regression, standard errors are clustered at the flag level. Panel C reports sample averages of freight rates (or fixtures).

Panel A: Tanker sample

No. unique tankers	5,363			
	Sanction risk			
Tanker type	Low	Moderate	High	Total
Medium	1,579	423	306	2,308
Panamax	297	88	55	440
Aframax	533	295	280	1,108
Suezmax	309	179	167	655
VLCC	672	64	116	852
Total	3,390	1,049	924	5,363

Panel B: Tanker characteristics balance test

	Treated (High/Mod.-risk)	Control (Low-risk)	Diff.	t-value
Identity Change	0.169	0.165	0.003	[0.12]
Risky Flag	0.276	0.264	0.012	[0.24]
Ship-to-ship Transfer				
• anchor 10 km away from ports	2.652	2.733	-0.082	[-1.61]
• anchor 20 km away from ports	1.424	1.475	-0.051	[-0.67]
• draft change 10 km away from ports	1.620	1.668	-0.048	[-1.63]
• draft change 20 km away from ports	1.297	1.334	-0.037	[-0.98]
Irregular Trajectory				
• distance cutoff = 1 km	1.087	1.064	0.024	[0.40]
• distance cutoff = 2 km	1.426	1.414	0.012	[0.23]
• distance cutoff = 5 km	1.420	1.434	-0.014	[-0.27]
DBSCAN Outlier				
• distance parameter = 10 km	3.744	3.670	0.075	[1.33]
• distance parameter = 20 km km	3.023	2.946	0.077	[1.31]
Dark Activity				
• at least 2-day dark period	0.883	0.872	0.010	[0.27]
• at least 3-day dark period	0.600	0.589	0.011	[0.31]

Panel C: Average fixture

	Average fixture in July 2023 (one month before the event)		
Full sample	Low risk	Moderate risk	High risk
165.3	129.8	135.2	126.8

Table 5: Difference-in-differences analysis of fixture changes

This table reports the coefficients (i.e., average treatment effects on the treated, ATT) estimated with PSM-DiD and Abadie (2005)'s semiparametric DiD to examine the change in freight rates (or fixtures) for high/moderate-risk tankers after Windward's disclosure. The sample period is from Feb 2023 to Mar 2024. The post-period starts from Aug 2023. The propensity score is calculated by regressing the high/moderate-risk tanker indicator on pre-period tanker characteristics using ML methods: average outputs from a decision tree and neural networks. The variables used to construct the propensity score are shown in Table 4. In panel A, we match tankers within each tanker type and calculate weights based on propensity scores and a Gaussian kernel with a bandwidth of 0.01, 0.03, or 0.05. Then, we do DiD estimation in the matched sample, controlling for tanker fixed effects and time×tanker-type fixed effects. The standard errors in panel A are double clustered at the tanker and time×tanker-type levels. In panel B, we implement Abadie's method. We first collapse our sample into two cross-sections by calculating each tanker's pre- and post-period average fixtures and then use the Stata command "absdid", which needs to specify the dependent variable (i.e., fixture changes from pre- to post-period), treatment dummy (i.e., high/moderate-risk tanker indicator), and propensity score. Since Abadie's estimator is derived without directly accounting for macro trends through time fixed effects, we manually subtract the cross-sectional mean to account for time fixed effects or the cross-sectional mean by tanker type to account for time×tanker-type fixed effects. We trim the fixtures at the 1st and 99th percentiles each month to avoid the influence of outliers. The standard errors in panel B are as derived in Abadie (2005). *, **, *** denote significance at the 10%, 5%, and 1% level, respectively.

Panel A: PSM-DiD			
	Dependent variable: Fixtures		
	Bandwidth = 0.01	Bandwidth = 0.03	Bandwidth = 0.05
ATT for High Risk	-16.063***	-15.150**	-14.605**
	[-2.69]	[-2.39]	[-2.31]
Obs. (tanker-month)	5,378	5,574	5,574
ATT for Moderate Risk	-0.684	-1.197	-1.493
	[-0.17]	[-0.32]	[-0.41]
Obs. (tanker-month)	6,081	6,083	6,083
Time × Tanker Type FE	Yes	Yes	Yes
Tanker FE	Yes	Yes	Yes
Panel B: Abadie (2005) Semiparametric DiD			
	Raw fixtures	Fixtures demeaned by time	Fixtures demeaned by time × tanker type
ATT for High Risk	-18.417***	-20.709***	-13.749***
	[-3.56]	[-4.47]	[-3.16]
Obs. (tanker)	1,036	1,036	1,036
ATT for Moderate Risk	-3.736	-5.087	-0.507
	[-0.88]	[-1.40]	[-0.15]
Obs. (tanker)	1,190	1,190	1,190

Table 6: Difference-in-differences analysis of route changes

This table reports the coefficients (i.e., average treatment effects on the treated, ATT) estimated with PSM-DiD and Abadie (2005)'s semiparametric DiD to examine the change of the routes of high/moderate-risk tankers after Windward's disclosure. The dependent variable is a zero-one indicator of passing during each month within 12 nautical miles (i.e., the boundary of the territorial sea) of Iran, Russia, or Venezuela (only before 2024). The sample period is from Feb 2023 to Mar 2024. The post-period starts from Aug 2023. The propensity score is calculated by regressing the high/moderate-risk tanker indicator on pre-period tanker characteristics using ML methods. The variables used to construct the propensity score are shown in Table 4. In Panel A, we match tankers within each tanker type and calculate weights based on propensity scores and a Gaussian kernel with a bandwidth of 0.01, 0.03, or 0.05. Then, we do DiD estimation in the matched sample, controlling for tanker fixed effects and time×tanker-type fixed effects. The standard errors in panel A are double clustered at the tanker and time×tanker-type levels. In Panel B, we implement Abadie's method in the same way as mentioned in Table 5. The standard errors in panel B are as derived in Abadie (2005). *, **, *** denote significance at the 10%, 5%, and 1% level, respectively.

Panel A: PSM-DiD			
	Dependent variable: Showing AIS signals near sanctioned countries		
	Bandwidth = 0.01	Bandwidth = 0.03	Bandwidth = 0.05
ATT for High Risk	-0.073***	-0.072***	-0.071***
	[-3.69]	[-3.76]	[-3.79]
Obs. (tanker-month)	39,246	39,246	39,246
ATT for Moderate Risk	-0.016	-0.015	-0.015
	[-1.01]	[-1.02]	[-1.05]
Obs. (tanker-month)	43,820	43,820	43,820
Time × Tanker Type FE	Yes	Yes	Yes
Tanker FE	Yes	Yes	Yes
Panel B: Abadie (2005) Semiparametric DiD			
	Probability	Probability demeaned by time	Probability demeaned by time × tanker type
ATT for High Risk	-0.067***	-0.067***	-0.060***
	[-4.01]	[-4.03]	[-3.60]
Obs. (tanker)	2,543	2,543	2,543
ATT for Moderate Risk	-0.012	-0.012	-0.008
	[-0.97]	[-0.96]	[-0.62]
Obs. (tanker)	3,139	3,139	3,139

Table 7: Difference-in-differences analysis of ownership changes

This table reports the coefficients (i.e., average treatment effects on the treated, ATT) estimated with PSM-DiD and Abadie (2005)'s semiparametric DiD to examine the effect of Windward's disclosure on the probability of tanker owner changes (i.e., liquidity). Our proprietary tanker ownership data contains six snapshots: Dec 2022, Mar, Jul, Sep, Dec 2023, and Mar 2024. The dependent variable is a zero-one indicator of owner changes by comparing the owner names across two snapshots. Note that the time spans between consecutive snapshots are different, so we make them comparable by annualizing the owner change variable. The pre-period owner change includes Dec 2022 - Mar 2023 and Mar - Jul 2023. The post-period owner change includes Jul - Sep 2023, Sep - Dec 2023, and Dec 2023 - Mar 2024. The propensity score is calculated by regressing the high/moderate-risk tanker indicator on pre-period tanker characteristics using ML methods. The variables used to construct the propensity score are shown in Table 4. In panel A, we match tankers within each tanker type and calculate weights based on propensity scores and a Gaussian kernel with a bandwidth of 0.01, 0.03, or 0.05. Then, we do DiD estimation in the matched sample, controlling for tanker fixed effects and time×tanker-type fixed effects. The standard errors in panel A are double clustered at the tanker and time×tanker-type levels. In Panel B, we implement Abadie's method in the same way as mentioned in Table 5. The standard errors in panel B are as derived in Abadie (2005). *, **, *** denote significance at the 10%, 5%, and 1% level, respectively.

Panel A: PSM-DiD			
	Dependent variable: Owner change		
	Bandwidth = 0.01	Bandwidth = 0.03	Bandwidth = 0.05
ATT for High Risk	-0.183***	-0.169***	-0.164***
	[-3.25]	[-2.91]	[-2.87]
Obs. (tanker-month)	11,402	11,402	11,402
ATT for Moderate Risk	-0.168***	-0.159**	-0.152**
	[-3.30]	[-2.79]	[-2.50]
Obs. (tanker-month)	11,776	11,776	11,776
Time × Tanker Type FE	Yes	Yes	Yes
Tanker FE	Yes	Yes	Yes
Panel B: Abadie (2005) Semiparametric DiD			
	Owner change	Owner change demeaned by time	Owner change demeaned by time × tanker type
ATT for High Risk	-0.185***	-0.184***	-0.170***
	[-3.30]	[-3.29]	[-3.04]
Obs. (tanker)	2,346	2,346	2,346
ATT for Moderate Risk	-0.168***	-0.167***	-0.146***
	[-3.20]	[-3.18]	[-2.77]
Obs. (tanker)	2,421	2,421	2,421

Table 8: Difference-in-differences analysis of changes of charterers

In this table, we examine whether U.S. or U.S.-allied companies tend to stop chartering high/moderate-risk tankers after Windward's disclosure. U.S.-allied companies are those from the UK, European Union, Australia, New Zealand, or Japan. The dependent variable is the number of tankers of a given risk type that each charterer employs in a month. The sample period is from Feb 2023 to Mar 2024. The post-period starts from Aug 2023. We regress the dependent variable on the indicators of U.S. and U.S.-allied charterers interacted with the post-period indicator controlling for charterer fixed effects and time fixed effects. The coefficients of the DiD estimation are reported. The standard errors are double clustered at the charterer and time level. *, **, *** denote significance at the 10%, 5%, and 1% level, respectively.

	Dependent Variable: No. tankers for each charterer-month		
	No. high-risk tankers	No. mod.-risk tankers	No. low-risk tankers
U.S. charterer $\times \mathbb{I}_{\{t\geq 0\}}$	-0.358*	-0.131	0.065
	[-1.91]	[-0.67]	[0.09]
U.S.-allied charterer $\times \mathbb{I}_{\{t\geq 0\}}$	0.013	-0.030	0.136
	[0.08]	[-0.12]	[0.31]
Obs. (charterer-month)	1,820	1,820	1,820
Time FE	Yes	Yes	Yes
Charterer FE	Yes	Yes	Yes

Table 9: Model parameters, moments matching, and implied quantities

This table presents the directly-calibrated model parameters in Panel A, the moments that we match in Panel B, the parameters derived from moments matching in Panel C, the imputed quantities in Panel D, and the model-implied effects due to price changes (i.e., we analyze the effects by fixing Qs in the models to pre-Windward levels) in Panel E. These results are discussed in Sections 5.2 and 5.3.

Panel A: Parameters directly calibrated

Variable	Calibration method	Symbol	Value
Discount factor	time-discounting (0.95) × depreciation	β	0.9
Proportion of Bad tankers	estimation in CRS (2024)	λ	0.213
Mean oil trade revenue (Clean exporters)	15.5 mln, normalized to one	r	1
Mean oil trade revenue (Rogue exporters)	4 dollars per barrel discount	r^R	0.794
Sanction penalty (Clean exporters)	4 mln, comparable to shipping costs	z	0.258
Sanction penalty (Rogue exporters)	50 mln, based on the median from actual cases of seized oil	z^R	3.226
Classification precision (pre-period)	implied by a pseudo-R^2 of 7.5% (public data's predictability in Figure 8)	ρ_{pre}	0.296
Classification precision (post-period)	implied by a pseudo-R^2 of 17.2% (Windward's predictability in Figure 8)	ρ_{post}	0.453

(continued)

Panel B: Matched moments

Moment	Model analog	Data	Model
Proportion of sanctioned tankers (average of pre- and post-Windward periods)	$\frac{Q^{pre}_{Sanc}+Q^{post}_{Sanc}}{2}$	0.5%	0.5%
Annualized volatility of low-risk tankers' fixtures before Windward's disclosure (divided by mean)	$\frac{\text{std}[p^{pre}_L]}{\mathbb{E}[p^{pre}_L]}$	0.77	0.77
Annualized volatility of high-risk tankers' fixtures before Windward's disclosure (divided by mean)	$\frac{\text{std}[p^{pre}_H]}{\mathbb{E}[p^{pre}_H]}$	0.81	0.83
Annualized volatility of top US charterers' revenue before Windward's disclosure (divided by mean)	$\frac{\text{std}[(Q^{pre}_{GL}+Q^{pre}_{GH}+Q^{pre}_{BL}+Q^{pre}_{BH})\tilde{r}]}{\mathbb{E}[(Q^{pre}_{GL}+Q^{pre}_{GH}+Q^{pre}_{BL}+Q^{pre}_{BH})\tilde{r}]}$	0.30	0.21
Disclosure effect on high-risk tankers' fixtures (DiD estimation in Table 5)	$\frac{\mathbb{E}[p^{post}-p^{pre}\vert\text{Post H}]}{\mathbb{E}[p^{pre}\vert\text{Post H}]}$	-0.13	-0.11
Disclosure effect on high-risk tankers' routes (DiD estimation in Table 6)	$\frac{\mathbb{E}[b^{post}-b^{pre}\vert\text{Post H}]}{\mathbb{E}[b^{pre}\vert\text{Post H}]}$	-0.17	-0.20

Panel C: Parameters derived from moments matching

Variable	Symbol	Value
Screening intensity (previous violation)	w_0	0.76%
Screening intensity (current violation)	δ_0	1.60%
Volatility of oil trade revenue	σ	0.365
Max operation cost when dealing with Rogue exporters	$\bar{c}$	0.872

Panel D: Imputed quantities (in mln)

Period	$\mathbb{E}(p_H)$	$\mathbb{E}(p_L)$	$\mathbb{E}(p^R_B)$	$\mathbb{E}(p^R_G)$	V_B	V_G
Pre-Windward	4.0	4.3	7.8	9.5	41.0	44.0
Post-Windward	3.7	4.0	7.5	9.9	36.9	41.0

Table 10: Model-implied effects due to price changes

This table reports the aggregate annual effect of the information shock on the four relevant parties – Rogue exporters, Clean exporters, Bad tankers, and Good tankers (in million dollars). We calculate the effects due to price changes by fixing quantities (i.e., the Qs in the model) to pre-Windward levels. For example, we report $\omega_G^{pre}\mathbb{E}(p_G^{R,pre}) + \omega_B^{pre}\mathbb{E}(p_B^{R,pre})$ as the per-tanker-trip cost for Rogue exporters in the pre-Windward period and $\omega_G^{pre}\mathbb{E}(p_G^{R,post}) + \omega_B^{pre}\mathbb{E}(p_B^{R,post})$ as the one in the post-Windward period, where $\omega_G^{pre} = [Q_{GL}^{R,pre} + Q_{GH}^{R,pre}]/[Q_{GL}^{R,pre} + Q_{GH}^{R,pre} + Q_{BL}^{R,pre} + Q_{BH}^{R,pre}]$ and $\omega_B^{pre} = 1 - \omega_G^{pre}$. We estimate the aggregate annual effects by multiplying per-tanker-trip costs/benefits by the number of tankers and 3.4 trips per year (the average number of trips per tanker from our fixtures data).

	Sanctions violators		Sanctions-compliant agents	
	R exporters (price paid)	B tankers (price received)	C exporters (price paid)	G tankers (price received)
Pre-Windward (per tanker-trip)	9.0	4.9	4.2	5.0
Post-Windward (per tanker-trip)	9.2	4.7	4.0	4.9
Percentage change	2.0%	-5.7%	-6.3%	-3.3%
Aggregate effect (annual, mln) (-/+ indicates loss/gain)	-780	-1,536	5,662	-3,346
Total aggregate effect (annual, mln) (-/+ indicates loss/gain)	-2,316		2,316	

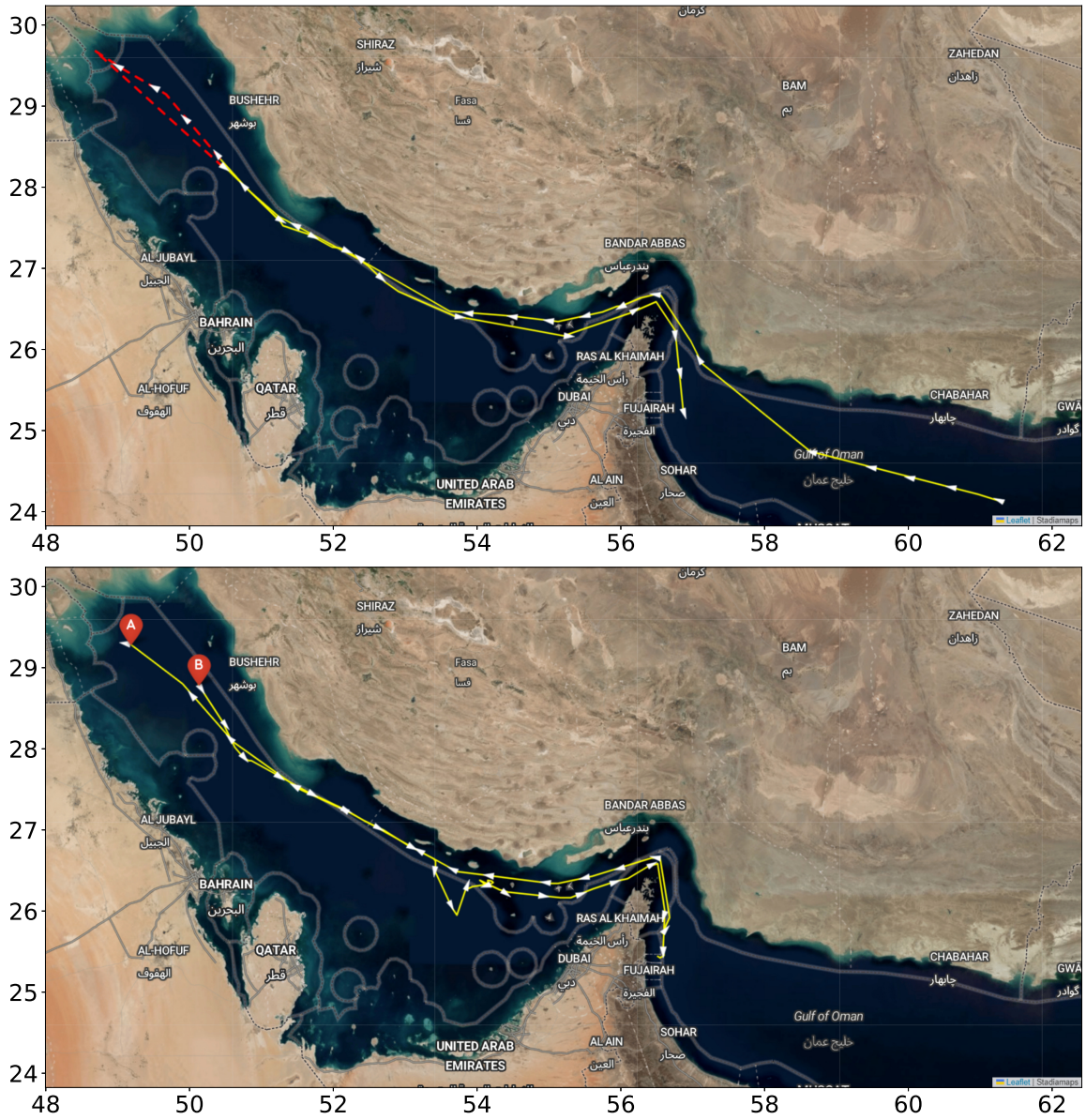


Figure 1: In this figure, solid and dashed lines show the trajectories of two tankers in the Persian Gulf, as given by their AIS signals (AIS is the Automatic Identification System for ship-tracking). White arrows indicate the direction of movement. The top plot shows a case of spoofing, i.e., at a time when the tanker was physically seen at an Iranian port (as per our dataset of sanctions-violating tankers), its AIS signals were showing it traveling in the northern part of the Gulf near Iraq (red dashed lines). Therefore, the tanker was manipulating its signals. The yellow solid lines show this tanker's path before and after its Iranian port visit. The bottom plot shows a case of dark activity, i.e., the tanker stopped emitting signals in a period when it was observed at an Iranian port (again as per our dataset). The last signal before it went dark is indicated by marker "A" and was emitted at 02:16:12 on 2021-01-23. The first signal after emerging from its dark period is indicated by marker "B", not too far from A and near the Iranian coast, and was emitted at 04:54:45 on 2021-01-27, ending a four-day dark period.

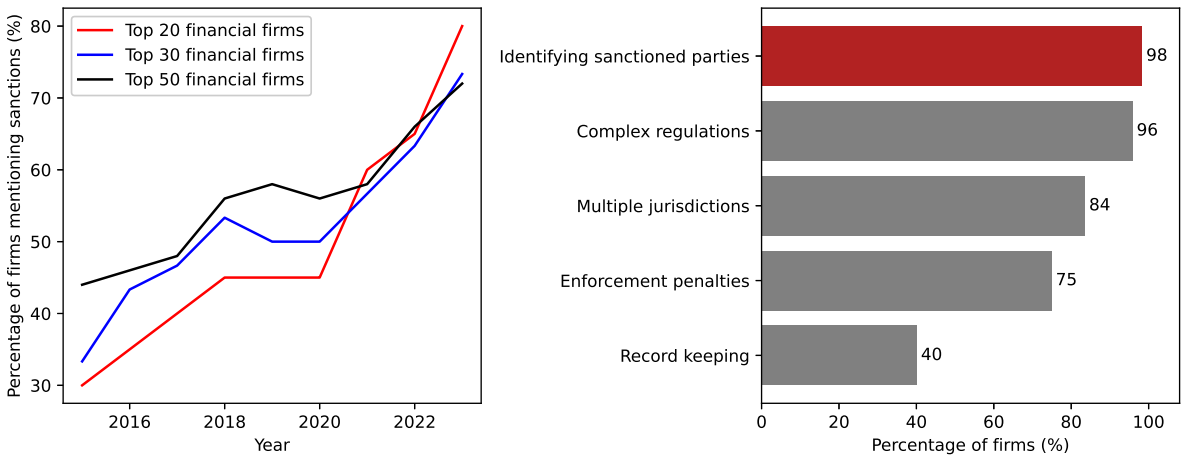


Figure 2: This figure shows the results from a large-language-model-based analysis of 10-K reports of the largest (by total assets) U.S. financial firms between 2015 and 2023. We first ask GPT-4 to identify the reports that mention sanctions enforced by OFAC, or sanctions-related risk, cost, or uncertainty. Then, we ask it to identify the reasons why the firms think sanctions compliance is challenging (further details are in Internet Appendix A). The left plot shows the percentage of firms among the 20, 30, and 50 largest financial firms, respectively, that mention economic sanctions-related risk (or cost or uncertainty) in their 10-K reports over the 2015-2023 period. The right plot shows the top five challenges in sanctions compliance for the 50 largest financial firms.

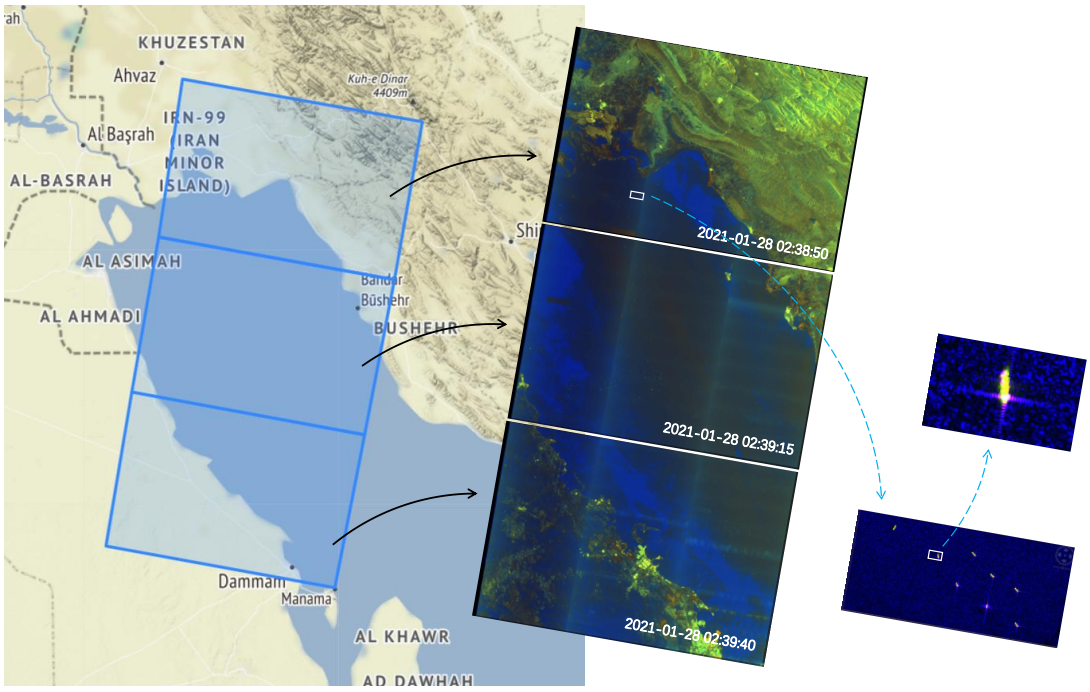


Figure 3: This figure shows examples of satellite images. These three images were taken a little after 2:30 a.m. on the 28th of January 2021 in the Persian Gulf area. Each image covers a rectangular sector, 250 kilometers wide. One insert on the right shows eight ships as little yellow objects at sea, while the other insert shows at the largest enlargement level one of these ships, with individual pixels clearly seen.

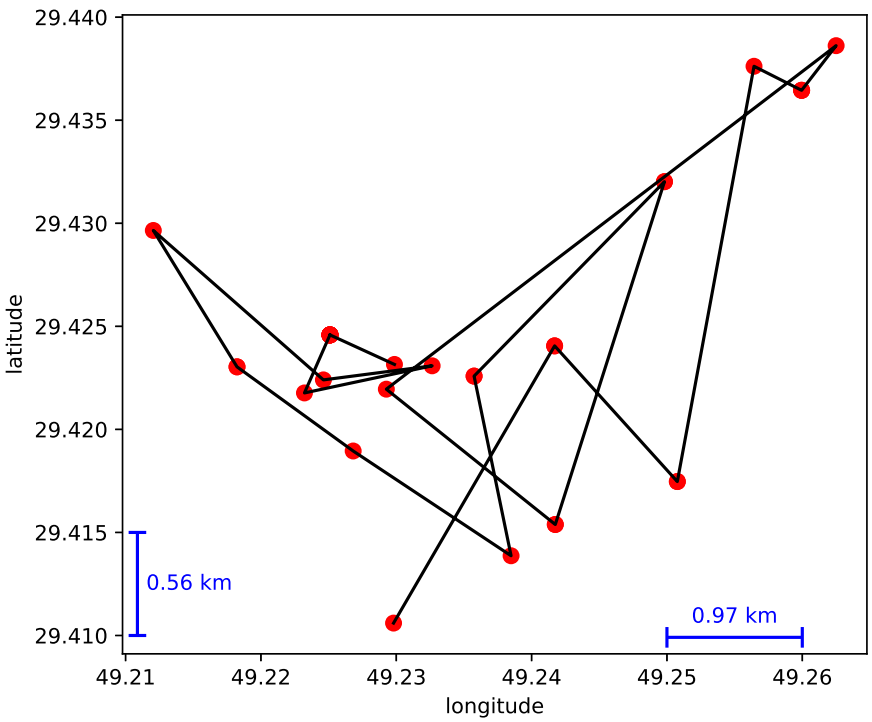


Figure 4: This figure illustrates an irregular trajectory. We define a trajectory as irregular if a tanker's total changes of course within a period are abnormally high. The axes show degrees of latitude and longitude. Red dots show the coordinates of each AIS signal emitted by each tanker.

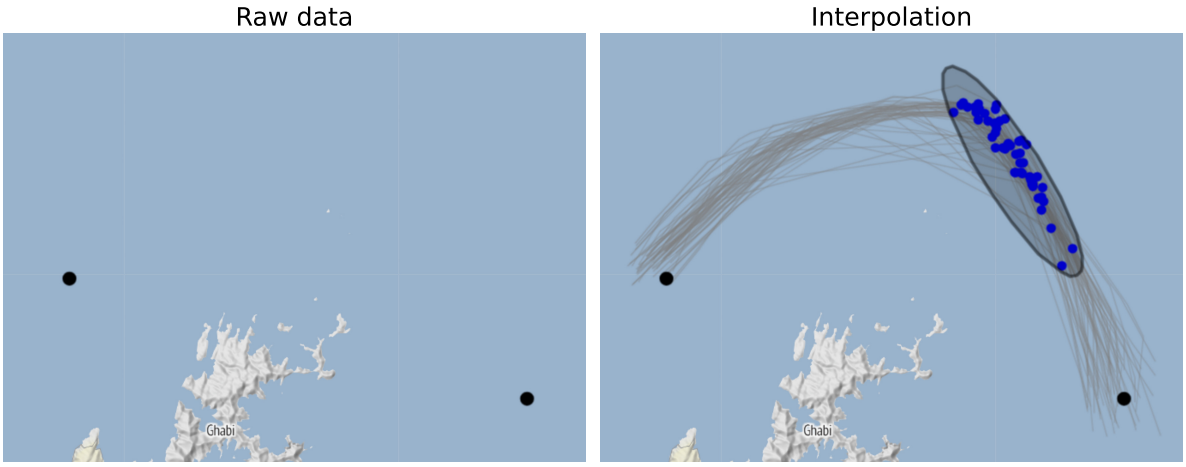


Figure 5: This figure illustrates the construction of the search area via interpolation. The black dots show the locations of a given tanker (call it A) at time t_0 (t_1) of the closest AIS signal before (after) the time t^I when a satellite could see the plotted area (the Strait of Hormuz). Let x_{t_0} and x_{t_1} denote tanker A's locations as per the AIS signals at t_0 and t_1. We require that t_0 and t_1 are within six hours of t^I. Tanker A was moving between these two times, and the distance between x_{t_0} and x_{t_1} is large. We determine the search area using the trajectories of other comparable tankers that have passed near both x_{t_0} and x_{t_1}, in the same direction and over a similar time interval, and hence at similar speed. We require a comparable tanker to have sent a signal within five kilometers from x_{t_0}, and then another one within the same distance from x_{t_1}, whereby the time it took to travel between these two signals is within 30 minutes of the difference $t_1 - t_0$. To increase the number of matches, we search for comparable tankers in the three-month period between December 2020 and February 2021. If multiple pairs of signals from a given tanker satisfy these matching criteria, we chose the pair with starting and ending locations closest to x_{t_0} and x_{t_1}. The grey lines in the figure show the trajectories of comparable tankers. Then we find by interpolation the location of each comparable tanker at its time corresponding to t^I for tanker A. These locations are shown as blue dots in the figure. Finally, we define the search area as the smallest ellipse that contains all of these interpolated locations (i.e., the blue dots), except those which are more than five standard deviations away from the center of the interpolation locations. We use ellipses to reduce the size of the search area.

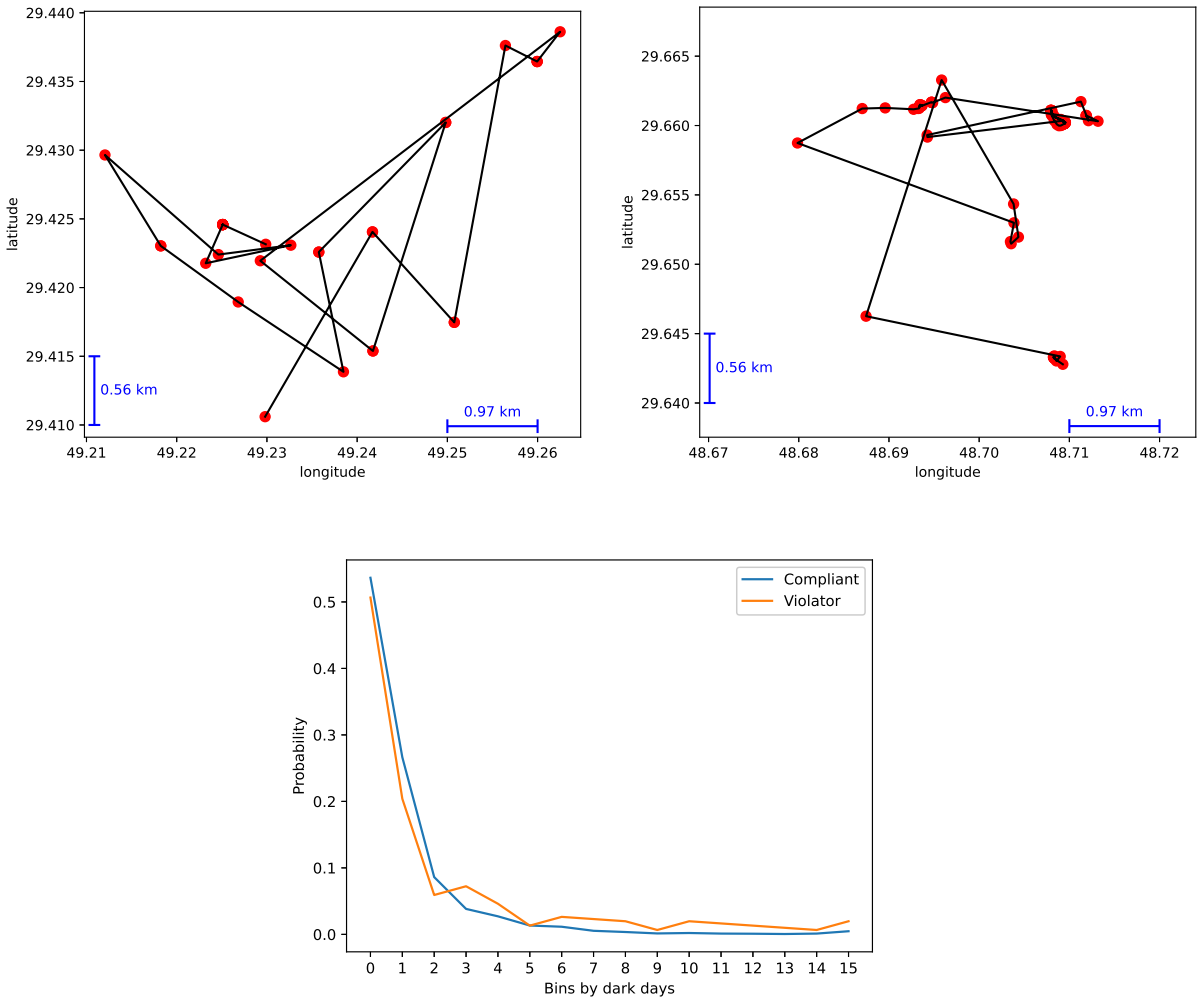


Figure 6: This figure shows irregular trajectories and distributions of dark days from sanctions-violating tankers and sanctions-compliant tankers. The top-left plot is for a sanctions-violating tanker from our proprietary dataset (it reproduces Figure 4 for direct comparability), and the top-right plot is for a sanctions-compliant tanker. The axes of these two plots show degrees of latitude and longitude, and the red dots show the coordinates of each AIS signal emitted by each tanker. The bottom plot shows the distribution of dark days in Jan 2021 for tankers entering the Persian Gulf conditional on the tanker going dark for at least half a day. Each bin in x-axis represents a range of dark time: bin 0 is from 12 hours to 24 hours; bin x ($x = 1, 2, \cdots, 14$) is from x to $x + 1$ days; bin 15 is equal or greater than 15 days. In this plot, the blue (yellow) line represents sanctions-compliant (sanctions-violating) tankers.

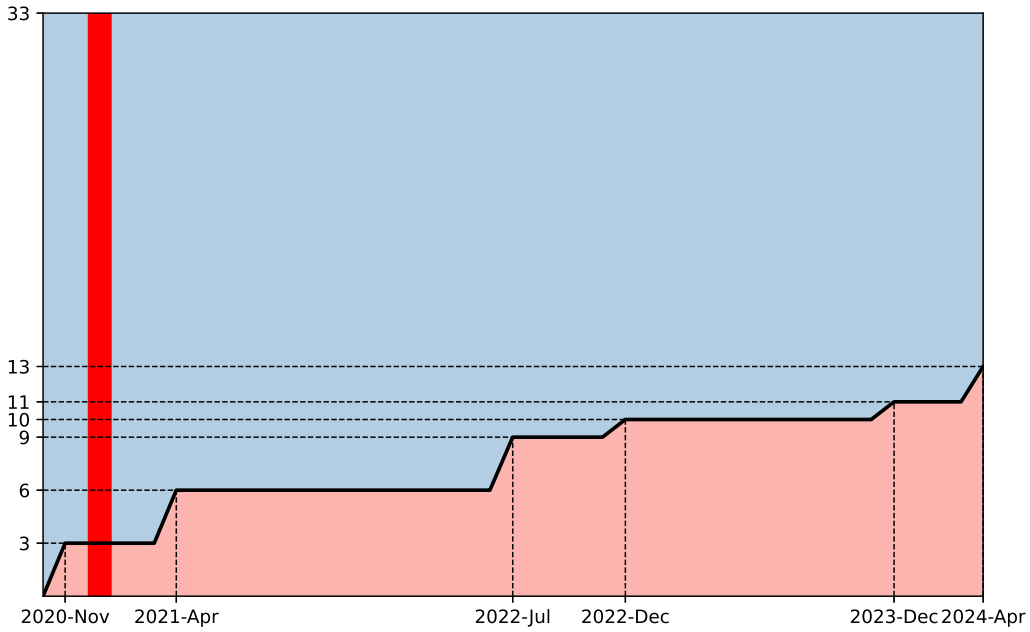


Figure 7: This figure illustrates the performance of UANI's "Ghost Armada" list of suspect sanctions-violating tankers. The solid black line shows at different dates the number of vessels in UANI's list that were also present in our dataset of 33 sanctions-violating tankers, that were observed physically at Iranian ports during January 2021. The line shifts up on dates when UANI's list is updated (shown on the x-axis). A red bar denotes the observation time of our dataset.

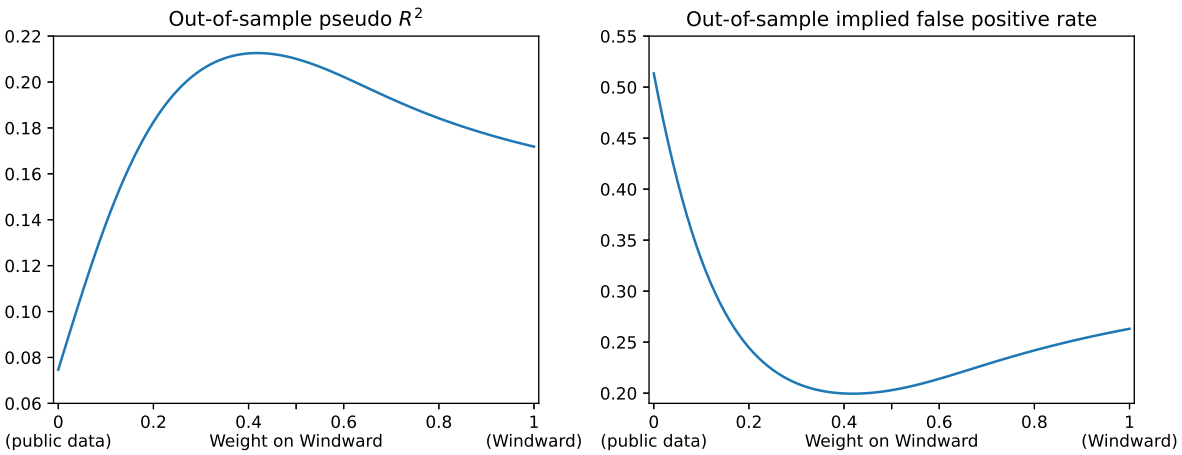


Figure 8: This figure compares the out-of-sample predictability of future sanctions achieved using weighted averages of Windward's and public-data-based predicted outcomes with various weights. We plot the corresponding out-of-sample performance measures against the weight for Windward's list. Windward's predicted outcome is a variable equal to one for high-/moderate-risk tankers and zero otherwise. For the public-data-based predicted outcome, we take the view of a third party predicting future sanctioned tankers at the end of July 2023 (i.e., right before Windward's disclosure). For our training sample, the predictors are constructed from data from August 2022 till January 2023 and the dependent variable is a zero-one indicator of tankers sanctioned between August 2022 and July 2023. The public data include AIS-based predictors and satellite-based predictors (as described in Section 3.3), and a zero-one predictor derived from the UANI list (as described in Section 3.6). We train ML models (decision trees and neural networks) with hyper-parameters optimally determined by 10-fold cross-validation. The public-data-based predicted outcome is the estimated probability of a tanker to be sanctioned after July 2023. The out-of-sample evaluation does not include tankers that were already sanctioned at the prediction time, i.e., the end of July 2023.

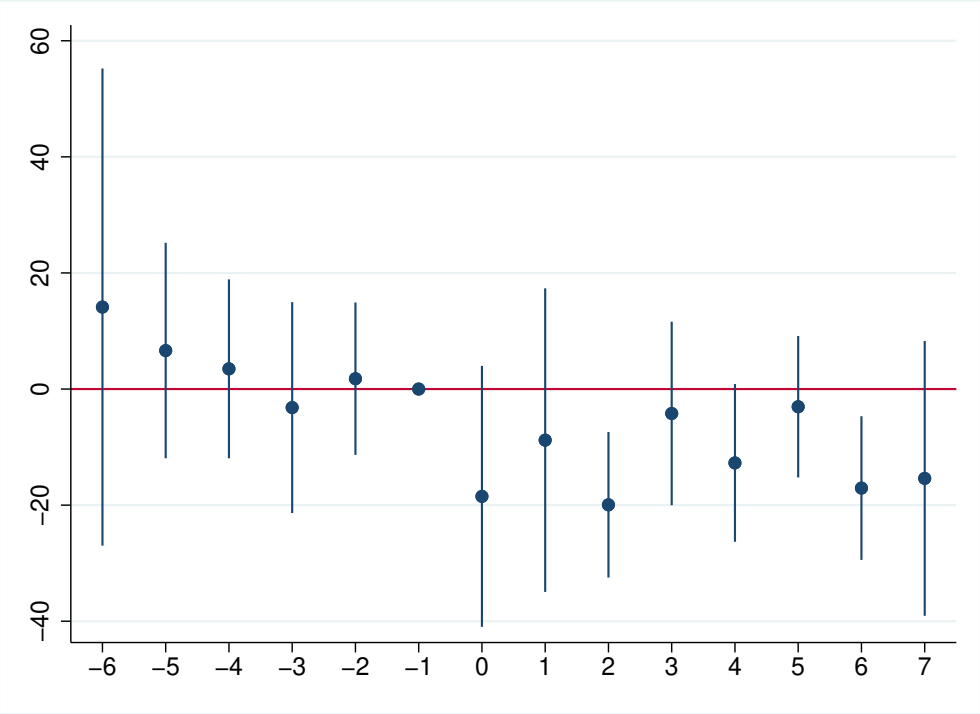


Figure 9: This figure plots the time series of coefficients (i.e., average treatment effects on the treated, ATT) estimated with PSM-DiD to examine the change in freight rates (or fixtures) for high-risk tankers only (i.e., we remove moderate-risk tankers from the sample). The sample period is from Feb 2023 to Mar 2024. Period 0 indicates the month of Windward's disclosure, i.e., Aug 2023. The propensity score is calculated by regressing the high/moderate-risk tanker indicator on pre-period tanker characteristics using ML methods: average outputs from a decision tree and neural networks. The variables used to construct the propensity score are shown in Table 4. We match tankers within each tanker type and calculate weights based on propensity scores and a Gaussian kernel with a bandwidth of 0.01. Then, we regress fixtures on indicators of high-risk tankers interacted by period dummies controlling for tanker fixed effects and time×tanker-type fixed effects. We trim the fixtures at the 1st and 99th percentiles each month to avoid the influence of outliers. Vertical bars represent 95% confidence intervals. The standard errors are double clustered at the tanker and time×tanker-type levels.

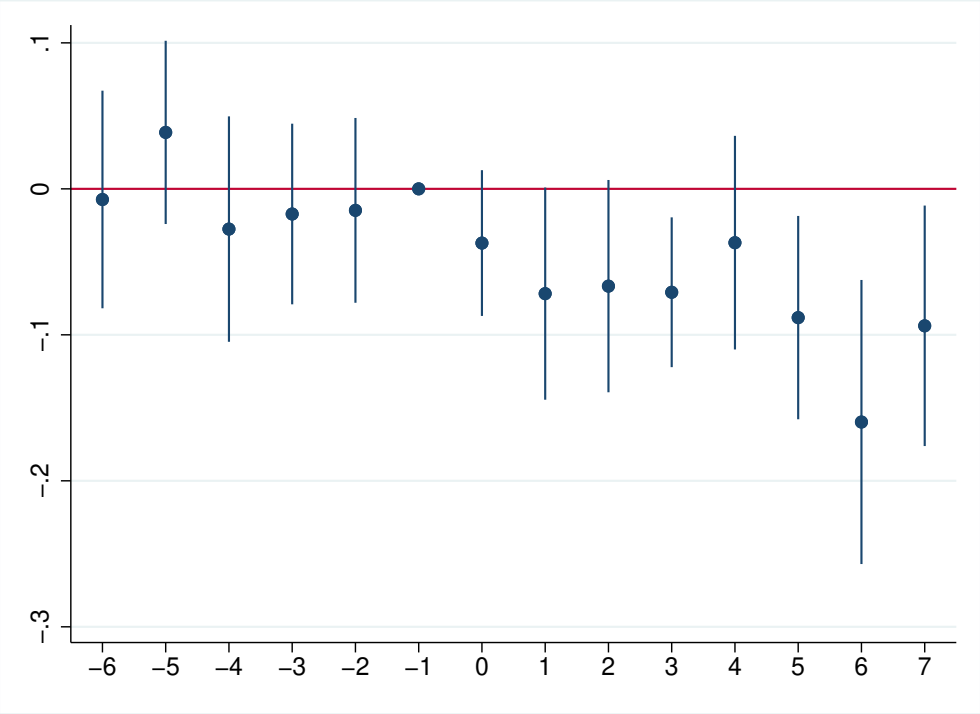


Figure 10: This figure plots the time series of coefficients (i.e., average treatment effects on the treated, ATT) estimated with PSM-DiD to examine the change of the routes of high-risk tankers only (i.e., we remove moderate-risk tankers from the sample). The dependent variable is a zero-one indicator of passing during each month within 12 nautical miles (i.e., the boundary of the territorial sea) of Iran, Russia, or Venezuela (only before 2024). The sample period is from Feb 2023 to Mar 2024. Period 0 indicates the month of Windward's disclosure, i.e., Aug 2023. The propensity score is calculated by regressing the high/moderate-risk tanker indicator on pre-period tanker characteristics using ML methods. The variables used to construct the propensity score are shown in Table 4. We match tankers within each tanker type and calculate weights based on propensity scores and a Gaussian kernel with a bandwidth of 0.01. Then, we regress the dependent variable on indicators of high-risk tankers interacted by period dummies controlling for tanker fixed effects and time×tanker-type fixed effects. Vertical bars represent 95% confidence intervals. The standard errors are double clustered at the tanker and time×tanker-type levels.

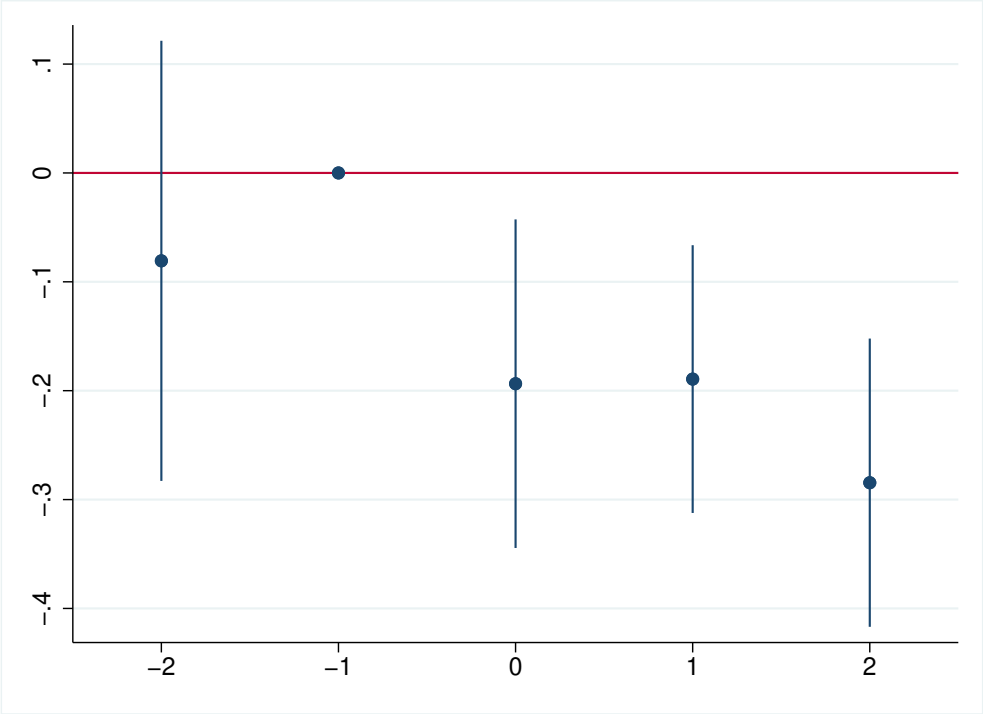


Figure 11: This figure plots the time series of coefficients (i.e., average treatment effects on the treated, ATT) estimated with PSM-DiD to examine the effect of Windward's disclosure on the probability of tanker owner changes for high-risk tankers only (i.e., we remove moderate-risk tankers from the sample). Our proprietary tanker ownership data contains six snapshots: Dec 2022, Mar, Jul, Sep, Dec 2023, and Mar 2024. The dependent variable is a zero-one indicator of owner changes by comparing the owner names across two snapshots. Since the time spans between consecutive snapshots are different, we make them comparable by annualizing the owner change variable. The pre-period includes Dec 2022 - Mar 2023 and Mar - Jul 2023. The post-period includes Jul - Sep 2023, Sep - Dec 2023, and Dec 2023 - Mar 2024. The propensity score is calculated by regressing the high/moderate-risk tanker indicator on pre-period tanker characteristics using ML methods. The variables used to construct the propensity score are shown in Table 4. We match tankers within each tanker type and calculate weights based on propensity scores and a Gaussian kernel with a bandwidth of 0.01. Then, we regress the dependent variable on indicators of high-risk tankers interacted by period dummies controlling for tanker fixed effects and time×tanker-type fixed effects. Vertical bars represent 95% confidence intervals. The standard errors are double clustered at the tanker and time×tanker-type levels.

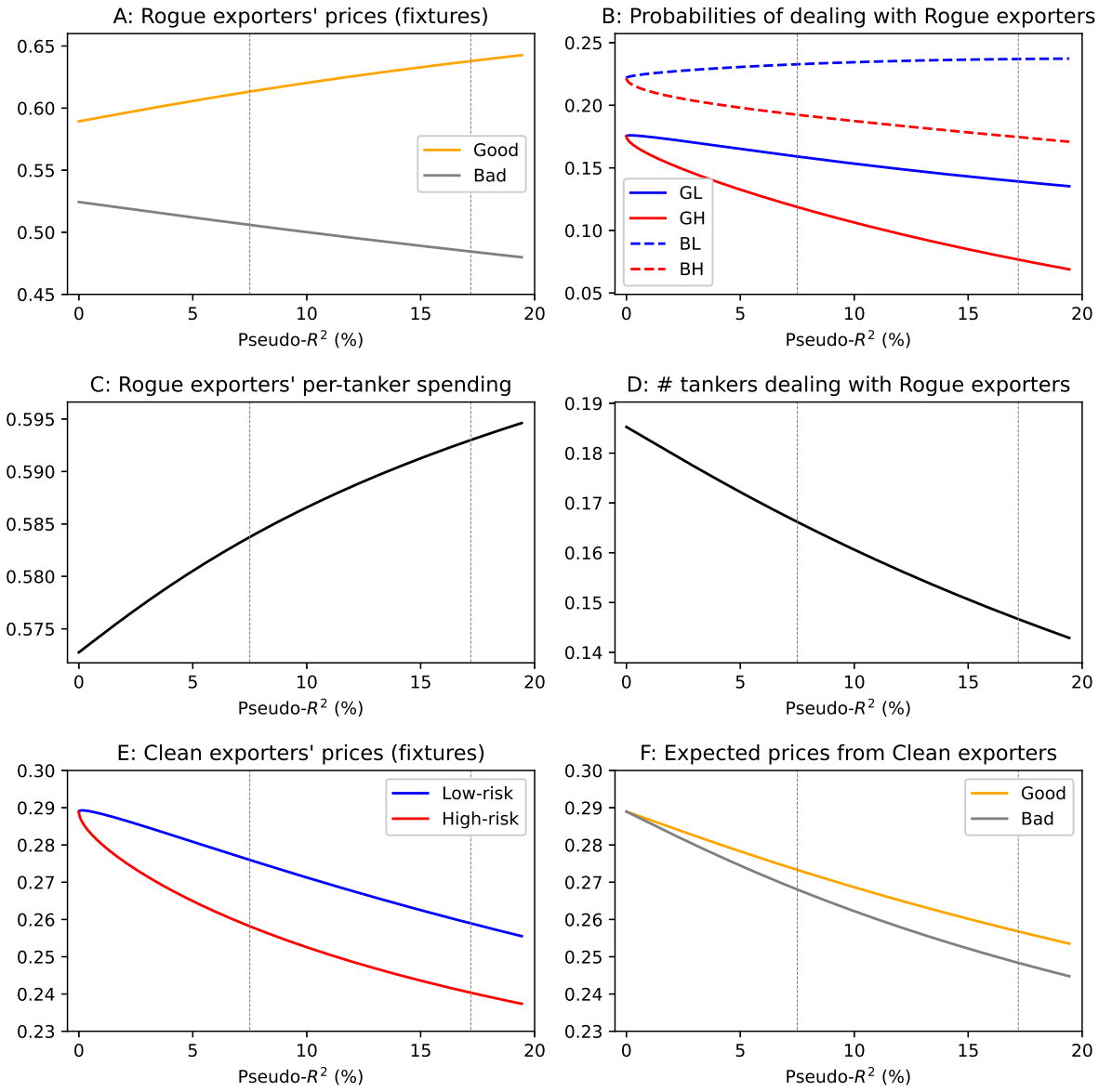


Figure 12: This figure presents the predictions from the calibrated model from Section 5. We solve the model by discretizing the aggregate shock $\tilde{\epsilon}$ on a grid. We plot the average model outcomes for the values of $\tilde{\epsilon}$ as functions of pseudo-R^2, which is backed out from the classification precision ρ in the model. In each subplot, two vertical dashed lines indicate the pre- and post-Windward precisions, which are 7.5% and 17.2%, respectively.